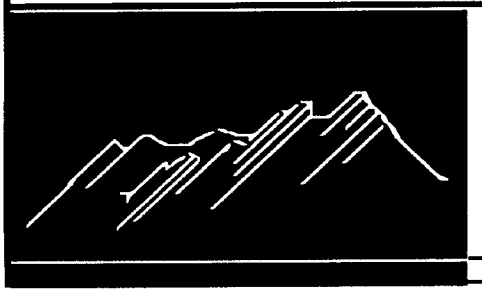
INSTITUTE OF COGNITIVE SCIENCE



Technical Report

University of Colorado, Boulder

# Practice Specificity In The Classic Stroop Color-Word Task

by

Deborah Marie Clawson

Institute of Cognitive Science
University of Colorado
Boulder, Colorado 80309-0344

ICS Technical Report #95-05

# Practice Specificity In The Classic Stroop Color-Word Task

Deborah Marie Clawson

Institue of Cognitive Science
University of Colorado
Boulder, CO 80309-0345

Abstract

Three experiments examined specificity of speed-up due to practice on the classic Stroop color-word task in which subjects are asked to name the color in which an incongruent color word is displayed. The results present a challenge to current views of automaticity and skilled performance that emphasize the importance of developing direct retrieval of correct responses from memory (e.g. Logan & Klapp, 1991). Specificity of practice effects has been used as support for those theories, but in this study specificity was found to the to-be-ignored words, implying that Stroop practice led to reduced memory retrieval of the words. It is suggested that theories of skilled performance that emphasize the role of direct memory retrieval be amended to include the role of decreased retrieval of incorrect responses as well as the role of enhanced retrieval of correct responses. In Experiment 1, Stroop practice led to both general improvement and improvement specific to the practiced stimuli; subjects trained on the Stroop task performed faster at posttest on the practiced color-word set than on a new color-word set . Subjects trained on simple color naming, however, were no faster on the Stroop task than were no-training control subjects, leading to greater interference than was found in the control condition. Experiment 2 replicated the Stroop-practice specificity found in Experiment 1 and further demonstrated that the specificity was largely due to the words in the practiced color-word set, with a trend of specificity to the practiced colors: Subjects were slower on a new color-word set than they were on the practiced color-word set; they also were slower on a stimulus set that used new words with practiced colors and, nonsignificantly, on a set that used new colors with practiced words. Experiment 3 showed that any specificity to color was not due to perceptual aspects of the practiced colors; there was no difference in posttest times on the practiced set and on a set wherein the colors had been slightly (but discriminably) changed both in chromaticity and luminance.

CONTENTS

# CHAPTER I
# INTRODUCTION

The question of whether improvements in task performance due to practice are specific or general has been of interest since at least the beginning of this century (Thorndike & Woodworth, 1901). In the past decade, the topic has enjoyed renewed interest, with a focus on specificity (Healy et al., in press; Logan & Klapp, 1991; Masson, 1986; Proctor & Healy, in press; Rickard & Bourne, in press; Rickard, Healy, & Bourne, in press; Whittlesea & Brooks, 1988), largely because of its implications for skilled performance and automaticity. This study examines specificity of practice improvements in one of the most studied phenomena of the last three decades, the Stroop effect (MacLeod, 1992; Stroop, 1935; for reviews of the literature, see Dyer, 1973; Jensen & Rohwer, 1966; MacLeod, 1991), because it poses a challenge to models of automatic performance supported by the specificity found in other domains. Although specificity to practiced colors in the Stroop task would be compatible with current views of the important role of direct memory retrieval in skilled performance, specificity to practiced words could not be associated with direct retrieval from memory and is therefore not compatible with those theories. Additional purposes of this study were to examine the pattern of reaction time improvements during practice on the Stroop task and to explore the implications of practice on simple color naming for Stroop performance.

Recent interest in the implications of practice specificity was in part kindled by Masson (1986) who found that practice improvements in reading inverted text were highly specific to the letters encountered during practice, to the typecase of those letters, and to clusters of those letters. After subjects who had practiced reading aloud words made up of mirror-image letters, they were much faster on those practiced words than on new words made up of unpracticed letters, on new words made up of practiced letters, on practiced words with new individual letter typecases, or on practiced words with new patterns of typecases. Subjects, although slower on the altered versions of practiced words than on the practiced stimuli themselves, were faster on all versions of the practiced words than they were on new words. Masson concluded that improved reading of inverted text depended on the development of instance-specific reading skills, and that any transfer was based on the new stimuli's sharing some features with practiced items. Whittlesea and Brooks (1988) found that experience copying pseudowords led to a word superiority effect identifying letters in those particular pseudowords. They took this as evidence for the role in letter identification of episodic experience with a specific stimulus.

Masson's interpretation of specificity as implicating instance-based learning was one later shared by Logan and Klapp (1991) in their study of alphabet arithmetic (e.g., A + 1 = B). Logan and Klapp further emphasized that automatic performance was due to memory for responses to specific instances encountered during practice. Such an "automaticity-as-memory theory" necessarily predicts specificity of practice effects. In examining specificity of practice effects in alphabet arithmetic, Logan and Klapp used as a performance measure the slope of reaction times as the number addend increased during subjects' verification trials on alphabet arithmetic problems; subjects in their verification task took longer to respond to problems in which the numbers were large than in which the numbers were small, but the effect of addend size decreased with practice. However, when subjects were transferred to a set of problems in which the numbers were the same as in practice, but the letters had not been in the practiced set, the effect of addend size increased again although not to the initial level. Verbal reports of counting to find the answers as opposed to remembering the answers increased markedly as well. In another experiment, Logan and Klapp further examined performance on variants of the trained stimuli, varied across two dimensions (thus comparing performance on new numbers with new letters, on new numbers with old letters, and on the practiced set), again finding specificity to the trained stimuli; performance on either set that included new numbers yielded greater slopes (that is, greater sensitivity to addend size) than was found on the practiced stimuli. They concluded that automaticity depends on retrieving the correct response directly from memory, not on better algorithm use.

Specificity of practice effects has also been found in studies of letter detection. In a recent review of research on skilled performance of letter detection, Proctor and Healy (in press) discuss the severely limited transfer they have found. They report a number of dimensions of specificity. Improvements due to practice detecting letters in random letter strings did not transfer to detecting letters in prose passages; intermediate performance on prose passages was found for subjects who had practiced on randomly arranged words. Improvements due to practice detecting letters in prose passages did not transfer to new target letters, even when the new target letter was present in the practiced high frequency word (e.g., *t* versus *h* on the word *the*), nor to passages that used new high frequency words with the practiced target letter (e.g. *the* versus *this* when detecting the target letter *t*). Proctor and Healy conclude that their findings of specificity are compatible with memory-based theories of skilled performance.

Most recently, Fendrich, Healy, and Bourne (1993) found practice specificity in mental arithmetic which led Rickard, Healy, and Bourne (in press) and Rickard and Bourne (in press), in expanded studies of mental arithmetic, to advance arguments for the importance for skilled performance of developing direct retrieval from

memory. Fendrich et al. found that subjects were faster on practiced single-digit multiplication problems than on operand-reversals of the practiced problems (e.g., 4 x 5 = _ and 5 x 4 = _) or on new problems. Rickard, Healy, and Bourne expanded that investigation, finding that after practice on single-digit multiplication and division problems subjects performed faster on practiced problems, on operand-reversals of practiced problems, and on versions of the practiced problems that used the same numbers and underlying operation (e.g. 4 x 5 = _ and _ ÷ 4 = 5) than they did on problems that differed from practiced problems in either numbers or operation (e.g. 4 x 5 = _ and 4 x 6 = _ or 4 x _ = 20, respectively). Rickard and Bourne report on a further expansion of these studies, examining specificity in an invented mathematical task ("pound arithmetic") to allow them to trace the course of early learning and asking subjects to report on some trials whether they had used the provided algorithm or direct memory retrieval in answering the problems. They found that subjects reported more direct memory retrievals and less use of the algorithm as practice progressed, with concomitant speed-up of responses, directly implicating a transition from algorithm-based performance to memory retrieval-based performance. Reaction time performance and subject reports on a set of posttests suggested that the transition to retrieval was only evident on practiced problems, not on new problems or on complementary versions of the practiced problems which changed the operation required. All of these studies viewed specificity of practice effects as evidence that improvement in performance after practice was largely due to newly developed or strengthened abilities to retrieve directly from memory the correct responses to practiced items.

Although there has been little research on the specificity of practice effects in the classic Stroop task, there have been some findings of improvement on the task with practice. In the Stroop effect, subjects show interference when naming the ink color of an incongruous color word. Beginning with Stroop (1935) a number of studies have shown that with practice, reaction times on the Stroop task or its variants diminished (Dulaney & Rogers, 1994; Flowers & Stoup, 1977; Harbeson, Krause, Kennedy, & Bittner, 1982; Roe, Wilsoncroft, & Griffiths, 1980; Shor, Hatch, Hudson, Landrigan, & Shaffer, 1972; White, 1978), although the interference was not eradicated. Connor, Franzen, and Sharp (1988) examined the effects of combined practice on all three Stroop related tasks (color naming, reading, and Stroop) and found that performance improved on both color-naming and Stroop tasks. None of these studies, however, examined the pattern of learning to determine whether reaction times across practice blocks approximated a power function, a pattern that has been called "the power law of practice" because of its ubiquity (see Newell & Rosenbloom, 1981). This study will examine the pattern of reaction times during practice both for the Stroop task and for simple color naming.

Despite the research on practice effects in the Stroop effect there has been little research on the extent to which such practice, either on the color-word interference task or on a simple color-naming task, is specific to the particular colors or words used during training. Specificity of training has been explored in two studies using nonstandard versions of the Stroop task. In a digit counting task, Reisberg, Baron, and Kemler (1980) trained subjects to ignore a pair of digits (e.g., 2 and 4) and found that this training did not transfer perfectly to ignoring other digits (e.g., the digits 1 and 3) nor did it transfer to ignoring homophonic words (e.g., *to* and *for*). Some transfer was obtained, however, to the task of ignoring the digits printed as words (e.g., *two* and *four*).

Ménard-Buteau and Cavanagh (1984) used a Stroop task with incongruously colored objects. Subjects practiced naming the ink color of a word representing an incongruously colored object (e.g., the word *carrot* printed in green ink). They found that this training did not transfer to a version of their task with drawings of the objects rather than words.

Clawson, King, Healy, and Ericsson (in press; Healy et al., in press) examined the effects of practice on the classic Stroop effect, providing four subjects with 12 sessions of practice either on the Stroop task itself or on simple color naming. Orthographic manipulations in tests, before and after training, provided an indication of specificity to word form. Another measure of specificity was provided by the use of two different color sets. Although the trained subjects were presented with only one color set during their practice, all subjects were tested on both color sets. In examining the effects of practice on Stroop performance, Clawson et al. found that there was significant improvement from pretest to posttest on the Stroop task, but that improvement did not depend on training condition -- even the control group, which had no practice, improved. This result suggested that the lion's share of the improvement was due to learning on the pretest given to all subjects. This finding agreed with Stroop's (1935) data showing the greatest improvement after the first session of practice. Sacks, Clark, Pols, and Geffen (1991) also reported improvement only between their first two blocks of training, after which performance was asymptotic. The present study therefore does not include a pretest, in order to allow analysis of practice improvement from the beginning of training.

The improvement after Stroop training observed by Clawson et al. (in press) was found to be specific to the practiced color-word set; Stroop-trained subjects had faster reaction times on Stroop tests using practiced stimuli than on Stroop tests using a new color-word set. There are two dimensions of the Stroop task to which the specificity could adhere: color and word. A third possibility is that practice is specific to the particular combinations of color and word that were practiced. As mentioned earlier, specificity of practice effects in the Stroop effect holds a potential challenge to previous views of skilled

performance as a product of direct retrieval from memory. Any specificity to colors or to color-word combinations could be explained as direct retrieval -- in both cases, it would be possible to have a Stroop stimulus trigger retrieval of the appropriate color name. Specificity to the words used in training, however, could not be so easily explained. Specificity to practiced words would imply that after practice a Stroop stimulus would be less likely rather than more likely to trigger retrieval of the spelled word.

In the Clawson et al. (in press) study, a lack of effect of orthographic manipulations at test suggested that there may not be specificity to the word forms encountered during training; after practice on Stroop stimuli using lowercase letters in the words (e.g. *red*), subjects performed equally fast on those stimuli and on orthographic manipulations in which the letters were uppercase (*RED*) or bracketed by asterisks (**r*e*d**). It should be noted, however, that not all word form manipulations are ineffective in reducing Stroop interference; using an unusual definition of Stroop interference, Melara and Mounts (1993) found that making the words much more difficult to read (by using small text or padding the ends of the words with pound signs to equate them in length) reduced interference. That is, they found that making the words more difficult to read reduced the difference in reaction times between performance on the Stroop task and performance on a color-naming task in which the colors and words were congruent rather than incongruent (e.g. the word *red* in red ink). The possibility remains that there are specific practice effects for the semantic meanings of the to-be-ignored aspect (in this case, the word) as suggested by Reisberg et al. (1980).

In item analyses of practice reaction times in the Clawson et al. (in press) study, there was some evidence of specificity to colors, to words, and to color-word combinations. Evidence of practice specificity to the particular color-word combinations encountered during practice was also found in an experiment by Musen and Squire (1993); reaction times improved during Stroop practice with a color-word set in which each color was paired with only one word, but then increased when the practiced colors and practiced words were paired differently, although not to the level of initial performance on the original set.

Two major theories explaining the Stroop effect are relative speed of processing theories and automaticity theories (see MacLeod, 1991, for a thorough discussion of theoretical accounts of the Stroop effect). Based on the observation that adults read color words faster than they can name colors (Brown, 1915), the relative speed of processing explanation envisions Stroop processing as similar to a horse race: Reading and color-naming compete, with reading as the faster process. The resulting response competition causes interference. According to automaticity theories, the interference is caused not by a difference in processing speed but rather by a difference in automaticity between reading and color-naming.

Reading, being more automatic, requires less attention and may be irresistible. According to these theories, the more automatic process, reading, interferes with the lesser, color-naming (MacLeod & Dunbar, 1988). Both the automaticity theory of the Stroop effect and the horse-race theory predict that training on simple color naming, by increasing either the speed of color naming or its degree of automaticity, should improve Stroop performance.

In the Clawson et al. (in press) study there were no significant differences in performance, other than the Stroop group's specificity, after practice among groups trained on the Stroop task, trained on simple color-naming, or not trained. Therefore, it was unclear what the findings implied about the underlying cause of Stroop interference. There was improvement between pretest and posttest, improvement that differed between the different tests, with greater improvement on the tests that required subjects to name colors (simple color naming and Stroop) than on the tests that required subjects to read words (simple word reading and reverse Stroop). On one hand, both the relative speed of processing and the automaticity theories suggest that if simple color-naming practice results in improved color naming, then it should improve Stroop performance as well, and there was improvement on the two color-naming tests relative to the two reading tests. However, because this differential improvement was not influenced by training condition, it may be the case that training did not contribute enough improvement beyond the learning on the pretest to make a noticeable difference. The first experiment in the present study therefore reexamines the effects of simple color-naming practice on color-word interference in an effort to disentangle the previous findings.

The goals of this study were pursued in three related experiments on practice specificity in the Stroop task. The first experiment examined three aspects of Stroop performance: the effects of simple color-naming practice on Stroop performance, the specificity of practice effects with practice on simple color naming or on the Stroop task itself, and the pattern of reaction times during practice on the Stroop task. The first experiment is thus an effort to replicate and extend the findings of Clawson et al. (in press). The second experiment further explores the roles of words and colors in practice specificity, training subjects on a set of Stroop stimuli then testing them on four variants of the set: the trained stimuli, trained colors with new words, new colors with trained words, and new colors with new words. Finally, the third experiment focuses on the source of any color specificity, by unconfounding the stimulus color from the response terms; subjects trained on a set of stimuli were then tested on a variant of the stimuli in which the colors were slightly different shades of the same colors, thus requiring the same trained responses.

# CHAPTER II
# EXPERIMENT 1: STROOP AND SIMPLE COLOR-NAMING PRACTICE AND THE SPECIFICITY OF PRACTICE EFFECTS TO COLOR-WORD SET

The main goal of the first experiment was to establish the Stroop task as relevant to theories of skilled performance which predict specificity of practice effects. Therefore, subjects were trained on a set of Stroop stimuli then tested both on the practiced stimuli and on an unpracticed set of stimuli. Based on the view of skilled performance that emphasizes the role of direct memory retrieval, it was expected that specificity of practice effects would be observed as faster reaction times on practiced than on unpracticed stimuli.

This experiment therefore aimed to replicate and expand the findings of Clawson et al. (in press) by overcoming the clouding of results by learning on the pretest. In this experiment, the absence of a pretest allowed for clearer analysis of improvement due to practice than was possible in the earlier study.

Based on Clawson et al. (in press), it was predicted that subjects would improve with practice on the Stroop task. It was further expected, based on the ubiquity of power-law improvement, that the practice reaction times would follow a power function. Practice on simple color naming was expected to improve later Stroop performance by lessening the difference in speed or automaticity between color naming and word reading.

It was further predicted that specificity to the practiced color-word set would be found for the Stroop-trained group, but not for group trained on simple color naming. Relatively little evidence of forgetting across a delay interval was expected on the basis of Clawson et al. (in press) who found no significant forgetting across a retention interval, on the basis of previous studies showing extremely good retention of procedural skills (Healy et al., 1992) and on the basis of previous findings that spacing of practice and summary feedback format lead to good retention (Schmidt & Bjork, 1992). As in the Clawson et al. (1993) study, specificity effects were predicted to persist across the retention interval as well.

## Method

### *Subjects, Design, and Apparatus*

Twenty-four students from the University of Colorado, all demonstrating normal color vision, participated in partial fulfillment of a class requirement. Subjects were assigned to training condition, practiced set, and test order on the basis of their time of arrival for the initial session according to a fixed rotation.

A mixed-factorial design was employed, with one between-subjects factor, training condition (Stroop training, lines training, no training control), and three within-subjects factors, test time (posttest,

retention test), test type (lines subtests or Stroop subtests), and set type (practiced set, unpracticed set), and with two counterbalanced factors, test order and practiced set.

A DTK Data-1000 personal computer with a Zenith Data Systems color monitor was employed for training and testing. A MEL (Schneider, 1988) Version 5.0 voice key-button box and an Electret microphone were used for measuring the subjects' verbal response latencies and for recording the experimenter's indications of response accuracy.

*Materials*

Two types of stimuli were used: Stroop stimuli and lines stimuli. Each Stroop stimulus was a color word displayed in an incongruent color; each lines stimulus was a simple set of lines (in the configuration shown in Figure 1) displayed in a color. There were two types of training, lines training and Stroop training, each of which consisted of color-naming practice on only the indicated stimuli. A third group, the no-training group, received no color-naming practice. There were 240 trials in each training session, divided into 20 blocks of 12 trials each.

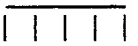


Figure 1. Configuration of lines used in the lines stimuli of Experiment 1. Note that although this figure is shown in black, the experimental stimuli were displayed in color.

The stimuli were presented in a pseudorandom order with the constraint that the stimuli possible for a given test were used equally often in each half block of six trials. For the Stroop training group, each color-word combination appeared once in each block of six trials; for the lines training group, each color appeared twice in each half block of six trials. Two versions of the training stimulus sequences, differing only in the order of the stimuli within a given half block of six trials, were used for these two sessions.

Test materials were identical for posttest and retention test, consisting of four subtests each of which was preceded by instructions: 24 Stroop trials on the practiced set, 24 trials on the unpracticed set, 24 lines trials on the practiced set, and 24 lines trials on the unpracticed set. As in training, the stimuli occurred in a pseudorandom order with the constraint that the possible stimuli were used equally often in each half block of six trials. The six Stroop stimuli used in a set consisted of all possible incongruent combinations of the three colors and words from that set. The three lines stimuli used in a set consisted of lines in the three colors of that set. All subjects were first tested on both sets in the Stroop test then tested on both sets in the lines test. Within the Stroop and lines tests, half of the subjects in each condition were first presented with the practiced-set trials then

presented with the unpracticed-set trials; the other half of the subjects completed the unpracticed-set trials first, followed by the practiced-set trials.

The stimuli from Set 1 were: pink, blue, and orange. The stimuli from Set 2 were: purple, green, and red. Half of the subjects in each condition were trained on Set 1; the other, on Set 2.

### *Procedure*

Before any testing, subjects were assigned to training condition, training color set, and test order, with two subjects in each combination of condition, set, and order. At the beginning of the session, all subjects were given a standardized color vision test, which took approximately five minutes (Dvorine Pseudo-Isochromatic Plates, National Research Council, 1981). Then all subjects were trained on using the voice key and keyboard. For control subjects, the end of voice key and keyboard training marked the end of the first session. For the Stroop-training and lines-training subjects, this training was followed by approximately 30 minutes of color-naming practice on their relevant stimuli, with a 5-minute break in the middle. The second session of training consisted of the voice key/keyboard warm-up and, for the Stroop and lines groups, another 30 minutes of practice with break, followed for all subjects by the posttest. For all subjects, the retention session, 28 days after the second training session, consisted only of the voice key/ keyboard training and the retention tests.

#### *Voice key and Keyboard Training*

Before both training sessions and both tests, subjects were asked to warm up with the voice key and keyboard. They read aloud 10 digits shown on the screen one at a time and were given feedback informing them whether their verbal response was loud enough for the voice key to register. If the response was not sufficiently loud, the following warning message was displayed: "Sorry, I could not hear you. Please answer more loudly next time." If the response was loud enough for the voice key to register, then the computer advanced directly to the self-paced prompt for the next digit. The procedure for advancing to the next trial and for responding were identical to the procedure used later in color-naming practice.

Throughout all trials, subjects were given the warning message whenever their vocal response was not registered by the voice key (but the experimenter's classification of the response as "correct" or "incorrect", as described below, was registered by the button box). Any response not registered by the voice key was discarded from the data analyses.

#### *Color-Naming Practice*

For the Stroop and lines groups, after voice key and keyboard training, the Stroop or lines trials were begun. Instructions were self-paced and were shown on the computer screen. The instructions described the stimuli to be presented along with the expected response, and they included a single example from the appropriate set. The instructions directed subjects to respond as quickly as possible, while

maintaining a correct-response percentage of at least 85%. The subject was further instructed that on blocks in which percent correct was at least 85%, the computer would note "personal best" average reaction times and "top" times compared to those of the other subjects. The procedure on a given trial was as follows: First, subjects viewed a screen with the instructions: "PRESS THE SPACE BAR TO CONTINUE." When the subject pressed the space bar, the screen became blank for 300 ms. Next the stimulus appeared on the screen, and the subject responded, which caused the stimulus to disappear. The experimenter, sitting behind the subject, indicated the accuracy of the subject's vocal response by pressing a "correct" or "incorrect" response key on the button box. The experimenter's response prompted the computer to display the instruction: "PRESS THE SPACE BAR TO CONTINUE."

A summary feedback of the percent of correct responses and the average correct reaction times was shown for 2,000 ms at the end of every block of 12 trials -- for example: "Your average correct reaction time was 530 milliseconds." Each time that a subject's average correct reaction time was a personal minimum (and accuracy was 85% or higher), the subject saw the phrase "Personal Best!!" displayed below the feedback message, and a video-game tune was played. Furthermore, if the subject's average correct reaction time was in the top five block averages up to that time by any of the subjects, the CRT displayed the phrase "Top time!!!" with another video-game tune. If the subject's accuracy was less than 85% correct in a block, then feedback was limited to, "Your percent correct was below 85."

*Tests*

At the beginning of both test sets, subjects were instructed that they would be playing four "computer games." The instructions further stated that although there would be no feedback during the games, "what's important is how FAST you can respond...Your answers also have to be correct at least 85% of the time, or the score doesn't count." Each of the four subtests (Stroop and lines on practiced and unpracticed sets) was preceded by instructions listing the colors (and, for the Stroop subtests, the words) that would be used in the game as well as giving an example stimulus and correct response. As with training, the test instructions and trials were self paced. Test trial procedure was identical to training procedure without summary feedback.

After completing the retention test, subjects were asked to fill out a questionnaire asking them about their experience in the experiment. Results of the questionnaire were used to develop strategy categories that were later used in classifying the questionnaire results of Experiment 3 and are not reported here. The contents of the questionnaire are described in detail in the method section of Experiment 3.

## Results and Discussion

Reaction times were measured from the time that the stimulus appeared on the computer screen until the subject initiated a verbal response. All analyses were completed on individuals' mean logarithm reaction times for correct answers only. Means provided in the figures are geometric means; that is, they are ten to the power of the mean logarithmic reaction times.

Accuracy was generally high and therefore was not analyzed. The proportion of trials on which subjects answered incorrectly or the voice key did not register the proper response was .05 across the test trials and .04 across the practice trials. Table 1 displays proportion correct on the tests and practice trials for the three training conditions.

Table 1
Proportion of trials responded to correctly in practice sessions 1 and 2 as well as the posttest and retention test by each of the three training groups, across subjects in each group.

| Training condition | Session 1 practice | Session 2 practice | Posttest | Retention test |
|---|---|---|---|---|
| Lines | .96 | .97 | .95 | .95 |
| Stroop | .95 | .96 | .97 | .96 |
| No | -- | -- | .93 | .94 |

### *Performance on Tests*

Examining performance on the posttest and retention test led to the conclusion that simple color-naming practice did not lend any advantage on later Stroop performance. It was further found that Stroop practice led both to improvements specific to the color-word set encountered during training and to general improvements on later Stroop tests.

The pattern of reaction times on tests using Stroop stimuli and line stimuli during the posttest and retention test by subjects in the three different training conditions is shown in Figure 2. To examine retention over the four-week delay, a four-way mixed analysis of variance (ANOVA) was conducted on reaction times for these tests. As would be expected, subjects were faster overall on the lines subtests than on the Stroop subtests, $F(1,21) = 286.51$, $MSe = .0016$,

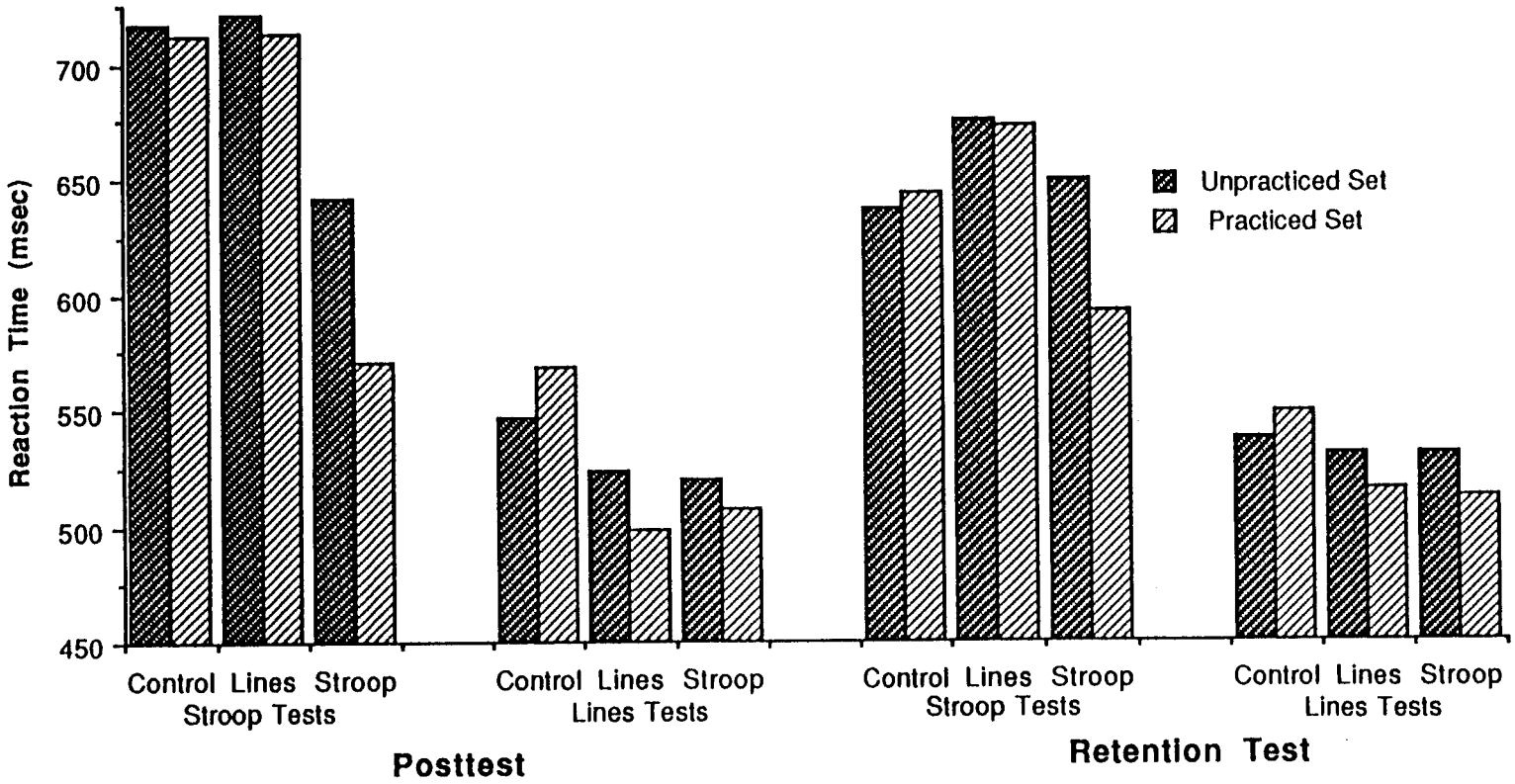



Figure 2. Mean reaction times of the three training groups (control, Lines, and Stroop) on the Stroop and Lines tests using practiced and unpracticed stimulus sets at the posttest and retention test.

$p < .001$. Although there were no overall differences in reaction times between the different groups ($F(2,21) = 1.04$, $MSe = .0206$, $p > .3$), the advantage on the lines subtest was different for the different groups, $F(2,21) = 8.07$, $MSe = .0016$, $p < .005$.

The most interesting aspect of the interaction between subtests and training conditions was that the difference in reaction times on the Stroop and lines posttests using the practiced set was greater for the subjects who had practiced with lines stimuli (mean difference between the logarithmic RTs = .156) and those who were in the control group ($M$ =.099) than for the subjects who had practiced on Stroop stimuli ($M$ = .051). Figure 3 illustrates this pattern with another view of the mean log reaction times on the Stroop and lines subtests using the practiced set at posttest and retention test. A post hoc analysis found a significant effect of training condition on the differences between reaction times on these tests, $F(2,21) = 11.05$, $MSe = .0020$, $p < .05$. In particular, the difference for the lines-trained group was significantly greater than that for the other two groups combined, $F(1,21) = 17.47$, $p < .001$, and the difference for the Stroop-trained group was significantly less than that of the control group, $F(1,21) = 4.58$, $p < .05$. (The post hoc tests were completed with a Scheffé adjustment using m = 3, m = 2, and m = 2, respectively.) Essentially, Stroop practice reduced the interference exhibited in the Stroop test (by speeding Stroop color naming even more than it speeded simple color naming), but lines practice increased the interference (by speeding simple color naming without speeding Stroop color naming).

On the retention test, the difference between times on the lines subtest and times on the Stroop subtest using practiced stimuli was again greater for the lines-trained group ($M$ = .12) than for the Stroop group ($M$ = .06) and control group ($M$ = .07) together, $F(1,21) = 10.70$, $MSe = .0012$, $p < .01$. At that test time, however, the difference was no longer significantly less for the Stroop-trained group than for the control group, $F(1,21) < 1$. (Again, these post hoc tests were both completed with Scheffé adjustments using m = 2.) Lines practice increased the interference by speeding simple color naming without speeding Stroop color naming even after the four-week retention interval.

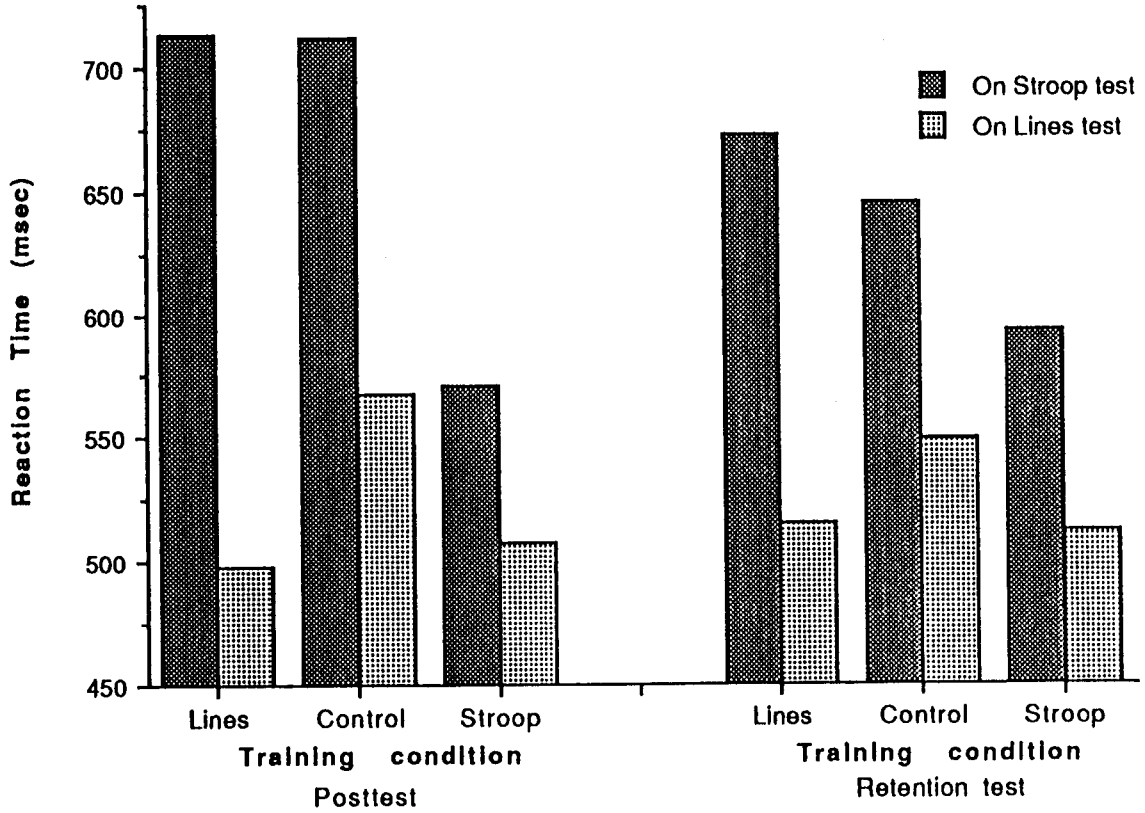



Figure 3. Experiment 1 Mean reaction times on the Stroop and Lines subtests using the practiced set at posttest and retention test for the three training conditions (Lines, Control, Stroop).

Subjects performed faster overall on the retention test than on the posttest, $F(1,21) = 4.53$, $MSe = .0010$, $p < .05$. This faster performance on the retention test was found only for the control and lines-trained groups, with the Stroop-trained group showing slightly faster performance on the posttest rather than the retention test, $F(2,21) = 5.80$, $MSe = .0010$, $p<.01$ for the interaction. Faster performance at retention further depended on the test type, with greater speed-up on the Stroop tests than on the lines tests, $F(1,21) = 8.34$, $MSe = .0007$, $p < .01$. This different speed-up on the different tests was not significantly affected by training condition, $F(2,21) = 3.08$, $MSe = .0007$, $p > .05$

The ANOVA revealed no effect of set type, across lines and Stroop subtests for all three training conditions ($F(1,21) = 2.99$, $MSe = .0019$, $p > .05$) nor any interactions of set type with the other factors ($F(2,21) = 2.72$, $MSe = .0019$, $p > .05$ for the interaction with training condition; $F(1,21) = 1.37$, $MSe = .0008$, $p> .2$ for the interaction with test times, and $F(2,21) = 2.57$, $MSe = .0008$, $p > .05$ for the three-way interaction with test time and training condition; $F < 1$ for all other interactions involving set type). However, specificity of practice effects was addressed in a series of planned comparisons. Notable specificity was found, but only for the Stroop-trained group. On the Stroop posttest, the Stroop-trained group performed much faster on the practiced color-word set than on the unpracticed set, $t(7) = 3.42$, $p < .05$. On that subtest at retention the Stroop group's advantage on the trained set over the untrained set was a nonsignificant trend, $t(7) = 1.88$, $p > .1$. The Stroop group did not show specificity to color-word set on the lines subtest at either test time (on the posttest, $|t(7)| < 1$; on the retention test, $t(7) = 1.80$, $p > .1$). The lines-trained group showed no specificity to color-word set on either the Stroop subtest ($|t(7)| < 1$ at both test times) or the lines subtest (on the posttest $t(7) = 1.43$, $p > .1$; on the retention test $t(7) = 1.53$, $p > .1$).

General learning was examined in another set of planned comparisons. On the Stroop subtest using the unpracticed color-word set, the Stroop-trained group out-performed the control group on the posttest ($t(15) = 2.17$, $p < .05$) although not on the retention test, $|t(15)| < 1$. The lines-trained group showed no similar transfer on the lines subtest using the unpracticed color-word set, $|t(15)| < 1$ at both test times.

The effect of simple color-naming practice on later Stroop performance was examined in a set of planned comparisons of performance by the three training groups on the Stroop posttest using the practiced color-word set. There was a striking lack of transfer from lines practice to the Stroop task. Lines-trained subjects demonstrated improvement on the practiced-set lines subtest by performing faster than did the control group, $F(1,14) = 6.18$, $p < .05$. However, on the practiced-set Stroop subtest, the lines-trained group performed no better than the control group, $|t(15)| < 1$. The Stroop-

trained group performed significantly faster on that test than did either the lines-trained group ($t(15) = -2.38, p < .05$) or the control group, $t(15) = -3.06, p < .01$. The finding of no advantage on the Stroop test after simple color-naming practice has an interesting parallel in work by Robert Proctor and Chen-Hui Lu (R. Proctor, personal communication, February 6, 1994) who examined the effects of practice on the Simon effect (Simon, 1990) a task similar to the Stroop task. In the Simon task, subjects must respond by a movement of the left or right hand, depending on the stimulus letter. The stimulus letter, however, can be placed in an incongruent location; for example, a letter indication "right hand" might appear to the subject's left. The incongruent location causes interference much like in the Stroop task. Proctor and Lu found that practice on the simple letter-hand task without incongruent locations lent no advantage on later tests with congruent locations, just as in this experiment practice on simple color naming lent no advantage on later Stroop performance .

*Performance during Practice*

Analyses of the pattern of reaction times during practice are reported in Appendix A. The patterns of reaction times for the lines group and the Stroop group were consistent with power law speed-up. The data, however, were noisy enough that other functions could not be ruled out as equally well fitted.

Reaction times during practice were also examined for item effects. Item analyses within each Stroop-trained subject in the study by Clawson et al. (in press) suggested advantages for specific words, colors, and color-word combinations. Similar item analyses were completed in this experiment, but across the four Stroop subjects trained on each practice set rather than within a single subject. These analyses would be expected to detect any advantages for particular colors, words, or combinations that were consistent across the subjects. Four item analyses were performed on practice data from the Stroop-trained subjects, one analysis each on the first and last 48 trials of practice for each practiced set, collapsed across all subjects on the color set. Each ANOVA examined three nonorthogonal effects -- word, color, and the interaction of the two. Because of the nonorthogonal nature of this design, due to the fact that no congruous stimuli were presented to the subjects, the degrees of freedom for the interaction were reduced to one. The only significant effect in any of the four analyses was on the analysis of the first 48 trials of practice for the group trained on Set 1 (pink, orange, blue); there was a significant effect of color, with reaction times fastest on the color blue and slowest on the color pink, $F(2,18) = 3.93, p < .05$. This one effect aside, there did not appear to be consistent advantages for particular practiced colors, words, or color-word combinations over others in the practiced set. Note that in their item analyses Clawson et al. (in press) found effects for words, colors, and their interaction; the difference can be explained by the nature of the analyses. In Clawson et al. only one subject was trained on each color-

word set, so the item analyses were each within a single subject; in this study the item analyses were each across four subjects. It is still possible that the individual subjects in this study showed advantages for particular words, colors, or word-color combinations, but those advantages were not consistent across all subjects.

# CHAPTER III
# EXPERIMENT 2: SPECIFICITY OF PRACTICE EFFECTS TO COLORS AND TO WORDS

The goal of the second experiment was to determine whether practice specificity in the Stroop effect reflected specificity to the words or to the colors used in practiced. It is the possibility of specificity to practiced words that would present a challenge to current views of skilled performance as memory retrieval. Although specificity to practiced colors would be compatible with direct retrieval from memory of the color name, specificity to practiced words could not be associated with direct retrieval from memory of a correct answer, because the words were always paired with two different colors during training. In this experiment, therefore, subjects practiced on a set of Stroop stimuli then were tested on both the practiced stimuli and on stimuli that used practiced colors but with unpracticed words. Slower performance on those stimuli than on the practiced stimuli would suggest specificity to words which could not be explained by enhanced direct retrieval from memory. Subjects were also tested on stimuli that used practiced words with unpracticed colors and stimuli that used unpracticed words and unpracticed colors in order to assess the more easily explained specificity to colors and to replicate the findings of Experiment 1, respectively.

In this experiment, all subjects were trained on the Stroop task. After training, all were tested under four different conditions that used different combinations of words and colors as stimuli: (a) trained words but combined with new colors (color-changed), (b) trained colors but combined with new words (word-changed), (c) trained colors and words (both-old), and (d) untrained colors and words (both-changed). On the basis of Experiment 1, it was predicted that reaction times for the both-changed subtest would be slower than for the both-old subtest. On the basis of the item analyses of Clawson et al. (in press) which suggested some specificity to colors and to words, it was predicted that reaction times would also be slower on the color-changed and word-changed subtests than on the both-old test. A second goal of Experiment 2 was to examine again the pattern of reaction times over the course of practice.

## Method

Sixteen students from the same subject pool as Experiment 1, all demonstrating normal color vision, participated in partial fulfillment of a class requirement. Subjects were assigned to training color set and test order on the basis of their time of arrival for testing according to a fixed rotation.

A mixed-factorial design was employed, with three within-subjects factors, word set (practiced, unpracticed), color set (practiced, unpracticed) and test time (posttest, retention test), with two

counterbalanced factors, training color set and test order. The combinations of the factors word set and color set created four subtests: both-changed, word-changed, color-changed, and both-old. Apparatus and training materials were identical to those of Experiment 1's Stroop training group.

All subtests were made up of four blocks, with each block presenting every possible incongruent color-word combination of the set once. The six stimuli used in the both-changed and both-old subtests consisted of all possible incongruent combinations of the three colors and three words in the set. Because the words and colors of the remaining two subtests were mismatched, preventing any congruent combinations, all nine possible color-word combinations were used. Thus, the both-changed and both-old subtests each consisted of 24 trials; the word-changed and color-changed subtests each consisted of 36 trials. The stimuli were presented in a pseudorandom order under the constraint that the possible stimuli were used equally often in each block of trials. There were four subtest orders, selected according to a Latin square, with four subjects tested in each order.

The color-vision test and the training procedure were identical to those used in the Stroop training condition of Experiment 1. As in Experiment 1 subjects received two sessions of practice, with a posttest at the end of the second session and a retention test four weeks later. Testing procedure also followed that of Experiment 1, with subjects reading self-paced instructions and completing trials for each of the four subtests according to their assigned test order.

As in Experiment 1, after completing the retention test, subjects were asked to fill out a questionnaire asking them about their experience in the experiment. Results of the questionnaire were used to develop strategy categories that were later used in classifying the questionnaire results of Experiment 3 and are not reported here.

## Results

As in Experiment 1, reaction times were measured from the time that the stimulus appeared on the computer screen until the subject initiated a verbal response. All analyses were completed on individuals' mean logarithm reaction times for correct answers only. Means provided in the text are logarithmic means, but means provided in the figures are geometric means; that is, they are ten to the power of the mean logarithmic reaction times.

Accuracy was generally high and therefore was not analyzed. The proportion of trials on which subjects answered incorrectly or the voice key did not register the proper response was .05 across the test trials and .05 across the practice trials. Table 2 displays proportion correct on the tests and practice trials across all subjects.

Table 2
Proportion of trials responded to correctly in Practice Sessions 1 and 2 as well as the posttest and retention test across all subjects.

| Session 1 practice | Session 2 practice | Posttest | Retention test |
|---|---|---|---|
| .93 | .96 | .94 | .96 |

*Performance on Tests*

Performance on the posttest and retention test replicated the finding of Experiment 1 that Stroop practice led in part to improvements specific to the color-word set encountered during training. Furthermore, on the posttest the specificity seems particularly sensitive to the words of the trained set, a specificity that persisted nonsignificantly over the four-week retention interval. On the posttest there was a nonsignificant trend for specificity to the trained colors as well.

The pattern of reaction times on the four subtests during the posttest and retention test is shown in Figure 4. To examine retention over the four-week delay, a three-way within-subjects ANOVA was conducted to examine the effects of changing the color, of changing the word, and of test time on reaction times. Across both test times, response times on the subtests that used the practiced words (both-old and color-changed) were faster than reaction times on the subtests that used unpracticed words (both-new and word-changed), $F(1,15) = 9.90$, $MSe = .0011$, $p < .051$, but the subtests using unpracticed colors did not lead to slower reaction times than did the subtests using practiced colors, $F(1,15) < 1$. The effect of changing the words was not dependent on which color set was used, $F(1,15) = 3.37$, $MSe = .0005$, $p > .05$ for the interaction.

As in Experiment 1, the ANOVA confirmed that subjects learned during the posttest; subjects performed faster overall on the retention test than they had on the posttest, $F(1,15) = 5.54$, $MSe = .0010$, $p < .05$. The change in reaction times from the posttest to the retention test was apparent on the subtests that used unpracticed words but not on the subtests that used practiced words, $F(1,15) = 9.68$, $MSe = .0004$, $p < .01$ for the interaction. The decrease in reaction times was not affected by which color set was used, nor by an interaction of color set and word set, $F(1,15) < 1$ for both.

Specificity of practice effects was further addressed in a series of planned comparisons. Replicating the findings of Experiment 1, notable specificity to the practiced color-word set was found on the posttest. Subjects performed much faster on the both-old subtest than on the both-changed subtest, $t(15) = 2.817$, $p < .05$. Indicating a strong role for the practiced words in the observed specificity, reaction times at posttest on the word-changed subtest were slower than on the both-old subtest, $t(15) = 4.40$, $p < .001$; indeed, as can be seen in Figure 10, posttest times on the word-changed subtest were not significantly

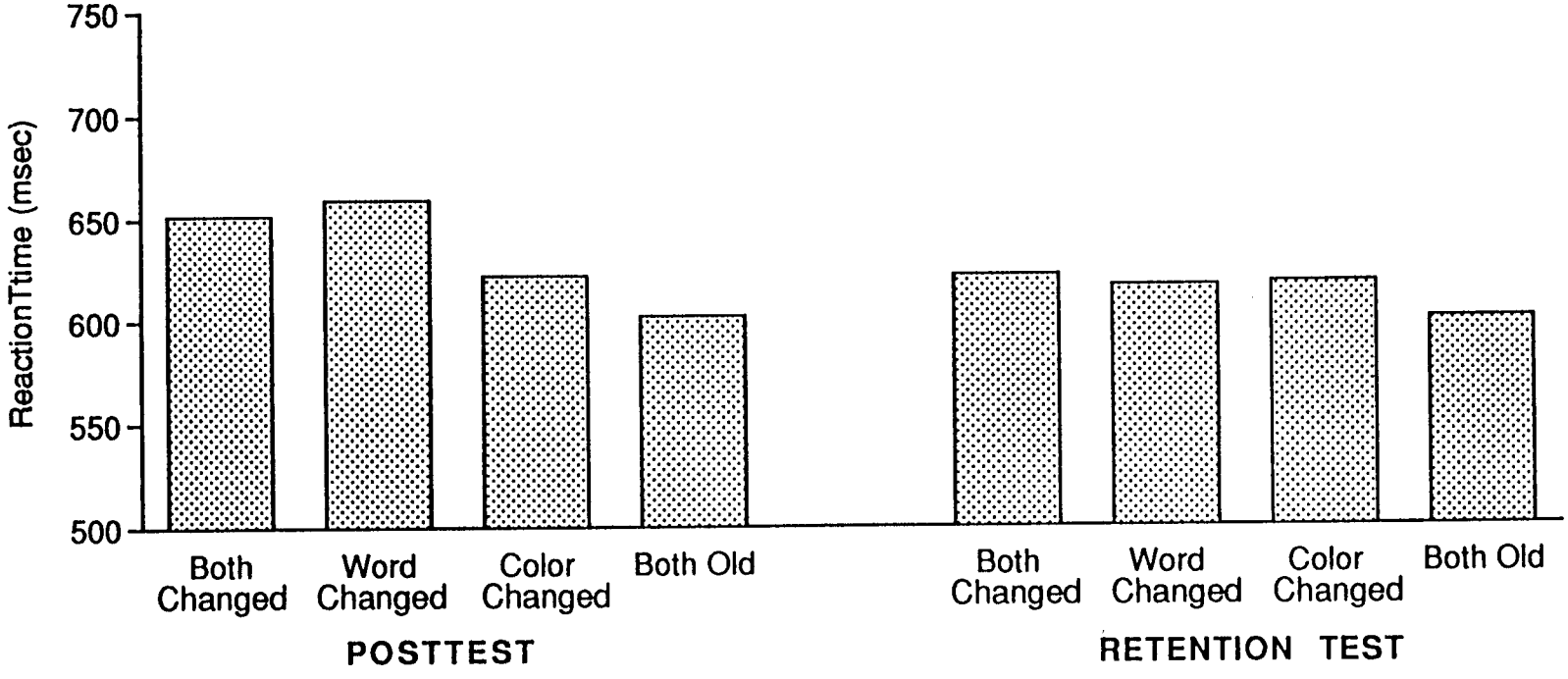



Figure 4. Experiment 2 mean reaction times on the four Stroop tests (both changed, word changed, color changed, and both old) at the posttest and the retention test.

different from those on the both-changed subtest, $|t(15)| < 1$. Implications for the role of the colors in Stroop specificity are less clear. Posttest times on the color-changed subtest were nonsignificantly slower than those on the both-old subtest ($t(15) = 1.36$, $p > .1$) and nonsignificantly faster than those on the both-new subtest, $t(15) = -1.88$, $p > .07$. As in Experiment 1, on the retention test the advantage of the both-old subtest over the other subtests was a nonsignificant trend ($t(15) = 1.33$, $p > .2$ comparing to the both-changed subtest; $t(15) = 1.36$, $p > .1$ comparing to the word-changed subtest; $t(15) = 1.10$, $p > .2$ comparing to the color-changed subtest).

*Performance during Practice*

Analyses of the pattern of reaction times during practice are reported in Appendix B. The pattern of reaction times was again consistent with power law speed-up. Reduced noise in the data allowed some distinction both in terms of estimates of fit and in terms of systematic deviations from the functions, suggesting that the power law accounted for the pattern of practice speed-up better than did simple exponential or linear functions.

# CHAPTER IV
# EXPERIMENT 3: SPECIFICITY OF PRACTICE EFFECTS TO PERCEPTUAL ASPECTS OF COLOR

Although determining the role of words in Stroop practice specificity has the greatest theoretical implications of these experiments, a complementary goal was to explore all the possible dimensions of specificity. Specificity of practice effects in the Stroop task could be composed of aspects both compatible with and incompatible with theories of skilled performance as direct memory retrieval. Experiment 1 established the existence of Stroop specificity. Experiment 2 determined that the to-be-ignored words played a major role in that specificity and that there may be a role for colors as well. Musen and Squire (1993) had found evidence for specificity at the level of color-word combinations. At the next level of analysis, both Clawson et al. (in press) and Musen and Squire (1993) found that it was not the perceptual aspects of words to which practice is specific, but perceptual aspects of the colors had not been examined. Therefore this experiment investigated the role of perceptual aspects of the colors.

In this experiment, all subjects were trained on the Stroop task. Then all were tested under three different test conditions all of which used the practiced words but each of which used different colors in the stimuli: (a) new shades of the trained colors with old words (shade-changed), (b) trained shades of the trained colors with old words (both-old), and (c) untrained colors with old words (color-changed). The new shades of trained colors were chosen such that the new and trained shades of the colors were identified (in a color-naming task completed by a separate group of subjects) with the same basic color names, but were discriminably different from each other. Specificity to perceptual aspects of the colors would be indicated by slower response times on the shade-changed subtest than on the both-old subtest.

A second goal of this experiment was to investigate whether subjects were aware of using any stimulus-specific strategies in their efforts to improve during practice. In responses to questionnaires administered at the ends of the previous experiments, subjects reported having used various strategies to attain faster performance. In this experiment, subjects' questionnaire answers were analyzed, using response categories based on answers in the previous experiments. Because it was clear from previous experiments that learning was largely specific to the practiced stimuli, the question of whether subjects were purposely using stimulus-specific strategies was of particular interest.

It was predicted on the basis of the Experiment 2 finding of a trend toward specificity to color that reaction times on the color-changed subtest would be slower than times on the both-old subtest. It

was further predicted that there would be some specificity to perceptual aspects of the colors as evidenced by slower reaction times on the shade-changed subtest than on the both-old subtest.

## Method

Twelve students from the same subject pool as Experiments 1 and 2, all demonstrating normal color vision, participated in partial fulfillment of a class requirement. Subjects were assigned to practiced set and to test order on the basis of their time of arrival for testing according to a fixed rotation.

A mixed-factorial design was employed, with two within-subjects factors, subtest (shade-changed, both-old, color-changed) and test time (posttest, retention test). Trained shade set and test order were counterbalanced across subjects; there were three test orders, selected according to a Latin square, with four subjects tested in each order. Apparatus and training materials were identical to those of Experiment 1's Stroop training group, except that the two practiced sets contained different shades of the same three colors, (red1, green1, purple1 and red2, green2, purple2). The different shades of the colors differed both in luminance and hue. Appendix C describes the perceptual qualities of the shades as well as the pilot study that determined that the alternate shades were identified with the same color name but were also discriminable. On the color-changed test, the colors were pink, orange, and purple.[1] All subtests were made up

---

[1] Whereas Experiments 1 and 2 counterbalanced the colors used in training, this experiment used only reds, greens, and purples during training, because of the technical difficulties of identifying alternate sets of color shades. It should be noted that a mixed four-way ANOVA of Experiment 2 test data revealed a significant interaction of the within-subjects factor color set (practiced, unpracticed) and the between-subjects factor practiced set, $F(1,14) = 15.92$, $MSe = .0009$, $p < .005$, such that across test times subjects who practiced on the red-green-purple set had longer times on the subtests that used unpracticed colors than on the subtests that used practiced colors, but subjects who practiced on the pink-orange-blue set showed little difference in speed between the two types of subtest. Therefore, Experiment 3 would be expected to yield greater specificity to colors than was found in Experiment 2 because in Experiment 3 only the red-green-purple set was used in practice, with the colors pink-orange-blue in the color-changed subtest, whereas Experiment 2 examined color specificity across both sets. It is important to note, therefore, that color specificity could be expected to be exaggerated in this experiment. Other findings in the four-way ANOVA reflected no effect of practiced set. Subjects in both practice groups had slower reaction times on the unpracticed-word subtests than on the practiced-word subtests (for the main effect, $F(1,14) = 10.29$, $MSe = .0011$, $p < .01$; for the interaction, $F(1,14) = 1.59$, $MSe = .0011$, $p > .2$). There was no significant main effect of practiced set, $F(1,14) < 1$. The main effect of the within-subjects factor test time and the interaction of test time and word set remained significant, $F(1,14) = 5.22$, $MSe = .0011$, $p < .05$, and $F(1,14) =$

of four blocks, with each block presenting every possible incongruent color-word combination of the set once. The six stimuli used in the both-old and shade-changed subtests consisted of all possible incongruent combinations of the three colors and three words in the set. In the color-changed subtest, because the words and colors were mismatched, preventing any congruent combinations, all nine possible color-word combinations were used. Thus, the shade-changed and both-old subtests each consisted of 24 trials, and the color-changed subtest consisted of 36 trials. The stimuli in a subtest occurred in a pseudorandom order under the constraint that the possible stimuli were used equally often in each block of trials. There were three subtest orders, selected according to a Latin square, with four subjects tested in each order.

The color-vision test and training procedure were identical to the Stroop training procedure of Experiments 1 and 2. Testing procedure also followed that of Experiments 1 and 2. As in the previous experiments, subjects received two sessions of training; at the end of the second session there was a posttest, and the retention test was administered two weeks (rather than four weeks) after the posttest. Again, subjects read self-paced instructions and completed trials for each of the three test conditions according to their assigned test order.

After the retention test, subjects were asked to fill out a questionnaire asking them about their experience in the experiment. Two questions asked subjects about any conscious strategies they may have used in performing the Stroop task: "If I asked you to tell somebody else how to do well at this game, what hints or advice could you tell them?" and "Did you use any particular strategies to help you do well on the game? What strategies? Did you ever think of any strategies that didn't work so well when you tried them? How about strategies that did work well? Please give us lots of details." Two more questions asked subjects their guess at the purpose of the experiment and what supposed condition of the experiment they had been in; the answers to these questions were examined for evidence that subjects had noticed the perceptual difference between the practiced and unpracticed shades.

## Results

As in Experiments 1 and 2, reaction times were measured from the time that the stimulus appeared on the computer screen until the subject initiated a verbal response. All analyses were completed on

---

9.81, $MSe$ = .0004, $p$ < .01, respectively. The main effect of the color set that was used in a subtest remained nonsignificant, $F(1,14)$ = 1.40, $MSe$ .0009, $p$ > .2, as did all other interactions: $F(1,14)$ = 3.16, $MSe$ = .0005, $p$ > .06 for the interaction of word set and color set; $F(1,14)$ = 1.20, $MSe$ = .0004, $p$ > .2 for the three-way interaction of test time, word set, and practiced set; $F(1,14)$ = 1.70, $MSe$ = .0006, $p$ > .2 for the four-way interaction of practiced set, test time, color set, and word set; $F(1,14)$ < 1 for all other interactions).

individuals' mean logarithm reaction times for correct answers only. Means provided in the figures are in ms; that is, they are ten to the power of the logarithmic mean.

Accuracy was generally high and therefore was not analyzed. The proportion of trials on which subjects answered incorrectly or the voice key did not register the proper response was .06 across the test trials and .06 across the practice trials. Table 3 displays proportion correct on the tests and practice trials across all subjects.

Table 3
Proportion of trials responded to correctly in practice sessions 1 and 2 as well as the posttest and retention test across all subjects.

| Session 1 practice | Session 2 practice | Posttest | Retention test |
|---|---|---|---|
| .93 | .95 | .94 | .94 |

*Performance on Tests*

Examining performance on the posttest and retention test replicated to a degree the finding in Experiment 2 that Stroop practice led in part to a specificity to the trained Set 2 colors, a specificity that persisted across the two-week retention interval. No specificity to the perceptual aspects of the colors was evident.

The pattern of reaction times on the three subtests during the posttest and retention test is shown in Figure 5. To examine retention over the two-week delay, a two-way within-subjects analysis of variance (ANOVA) was conducted to examine the effects of subtest and test time on reaction times. Neither test time nor subtest revealed a significant effect ($F(1,11) = 2.67$, $MSe = .0003$, $p > .1$ and $F(2,22) = 1.72$, $MSe = .0012$, $p > .2$, respectively), nor did their interaction ($F < 1$). In particular, the difference between the shade-changed and both-old subtests was not significant ($F(1,11) = <1$), nor was the difference between those two subtests together and the color-changed subtest, $F(1,11) = 2.54$, $MSe = .0016$, $p > .05$. Neither of the interactions of these factors with test time were significant, $F(1,11) < 1$ for both.

Specificity of practice effects was addressed in two planned comparisons. Just as in Experiment 2, subjects showed a trend of specificity to color; at both test times, reaction times were slower on the color-changed subtest than on the both-old subtest, nonsignificantly so on the posttest ($|t(11)| < 1$; on the retention test $t(11) = 2.21$, $p < .05$). On the shade-changed subtest, subjects were no slower than on the both-old subtest ($|t(11)| < 1$ at both test times), and no faster than on the both-changed subtest ($t(11) = -1.30$, $p > .2$ on the posttest; $t(11) = -1.11$, $p > .2$ on the retention test).

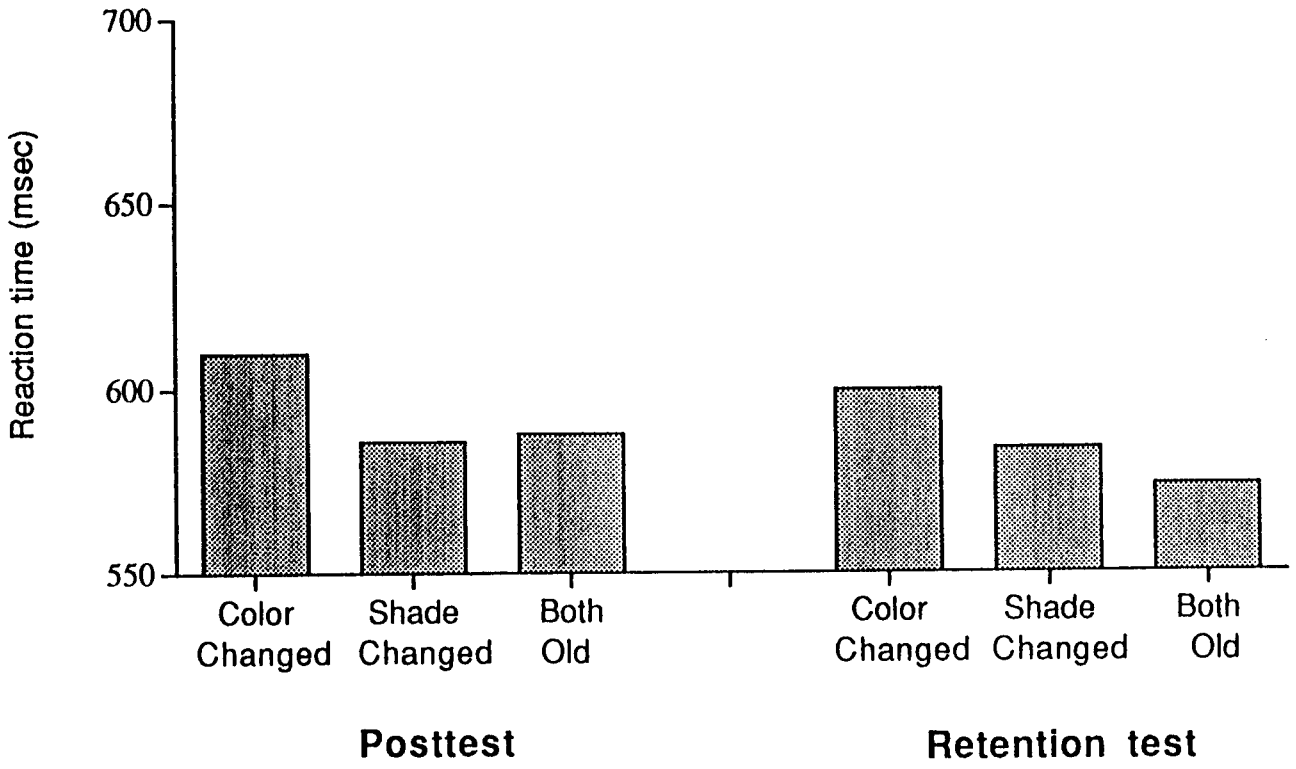



Figure 5. Experiment 3 mean reaction times on the color-changed, shade-changed, and both-old subtests of the posttest and retention test.

*Questionnaires*

On the basis of a task analysis and questionnaire responses from Experiments 1 and 2, subjects' strategies were classified into six categories by the Experimenter. The average number of strategies reported by a subject was 2.0, with the range 1 to 4. The strategy categories, along with the number of subjects reporting each, are listed in Table 4. It should be noted that although stimuli were presented in random order, the constraint that every stimulus be presented equally often in each block may have allowed some success with the strategy of anticipating the next color based on previous colors. Responses in the "Other" category were: memorizing what the colors were, relaxing the mouth, and pausing after mistakes. The blurring eyes strategy was reported by two (of the six listed) subjects to be an unsuccessful strategy; one of the subjects went further and explained the lack of success: "I thought that if I crossed my eyes I would only be able to see a blotch of color. However I could still make out 3 characters for R-E-D, 5 for green, & 6 for purple. So I was still able to read the word which always threw my game off." It is interesting to note that all of these strategies, save the "other" strategy of memorizing the colors, are general strategies that once learned should apply equally well to new stimulus sets and old. For example, subjects who have learned to blur their eyes on one color-word set should be able to blur them equally well on another set. Yet performance in Experiments 1 and 2 did show specificity to the trained set and trained words. Assuming that subjects in this experiment followed the same course of skill development during training (procedures in this experiment were identical to those of the previous experiments until the posttest), the finding of general strategies suggests that subjects may be unaware of any change in their performance that is specific to the practiced set.

Table 4
Number of subjects (N) reporting having used strategies from each category.

| **Strategy Category** | **N** |
|---|---|
| Concentrate on the colors | 7 |
| Blur eyes, cross eyes, focus eyes beyond the plane of the computer screen | 6 |
| Look at a some other location on the screen than the stimulus location | 3 |
| Anticipate the next color based on previous colors | 3 |
| Get into a rhythm of responding and do not pause between trials | 2 |
| Other | 3 |

On other questions, ones dealing with the purpose of the experiment, 4 of the 12 subjects mentioned that there had been two

different shades of red, green, and purple during the tests. For comparison, the same number of subjects (2 of whom had also reported the shade change) reported that there had been a new color set at test. Thus, subjects did notice the discriminable difference in shades without prompting.

# CHAPTER V
# GENERAL DISCUSSION

In these experiments, practice improvements on the classic Stroop task were found to be largely specific to the color-word set used in practice and particularly specific to the words in that set. To summarize the findings, in the first two experiments, subjects performed faster on the color-word set that they had practiced than on a new color-word set, replicating the findings of Clawson et al. (in press); there was some general Stroop improvement, as well. In the second experiment it was further found that subjects performed faster on the practiced set than on a set that used the practiced colors with new words. In Experiments 2 and 3 subjects' performance on a set using new colors with practiced words was nonsignificantly slower than on the practiced set itself, implying that Stroop practice may not be specific to the practiced colors; in Experiment 3 changing perceptual aspects of the colors had no effect on performance, leading to the conclusion that practice certainly was not specific to perceptual aspects of practiced colors.

The most important of these specificity findings is the specificity to practiced words. Such specificity could have important implications for the field of human learning. Similar strong improvement in task performance with specificity has been taken by Masson (1986), by Logan and Klapp (1991), and by Rickard and his associates (Rickard & Bourne, in press; Rickard et al., in press) as evidence that skilled performance -- and in Logan and Klapp's case, automatic processing -- is a product of direct retrieval from memory. Speed-up in performance of practiced tasks is due to the ability to recall the correct responses directly from memory without an intermediate algorithm or mnemonic.

The skill-as-retrieval view could easily account for either specificity to practiced color, in which seeing the color would result in direct retrieval of the color name, or specificity to color-word combinations, in which seeing a practiced word in a particular color would result in direct retrieval of the color name.

Specificity to the practiced words, however, is not so forthrightly explained. Dulaney and Rogers (1994) in studying the persistence of word suppression after practice, suggest that the word suppression had become automatic; in that sense, the findings of specificity to practiced words in this study might appear to be another line of evidence for automaticity as memory retrieval. Practice on the Stroop task led to an automatic skill that yielded specificity. However, this line of reasoning is complicated by an important aspect of the specificity to the to-be-ignored words. In the case of specificity to words, it would be necessary to postulate that practice at the Stroop task leads to not retrieving the word, in essence an explanation of

skilled performance as anti-retrieval. This implies that reduced memory retrieval of incorrect answers is as important to skilled performance of the Stroop task as is enhanced memory retrieval of correct answers to skilled performance of the tasks studied earlier.

Singley and Anderson (1989) identify the Stroop task as a rare example of procedural interference, in which doing one task is interfered with by a stronger response for doing another task. This could be taken as an indication that the Stroop task is outside the realm of normal skilled performance. However, Proctor and Lu's findings (personal communication, February 6, 1994) suggest that the Stroop task is a member of a class of similar problems in which skilled performance on one task interferes with performance of another task, an interference that can be overcome only by lessening the strength of the original response. That the Stroop task should be considered within the realm of skilled performance theories is also supported by the finding in the first two experiments that Stroop practice reaction times followed the power law of practice that is found in other skill domains. Thus, the role of anti-retrieval must be accommodated by theories of skilled performance and automaticity.

One alternative explanation for the specificity to words, other than the development of anti-retrieval for those words, can easily be ruled out. Verbal protocols of subjects practicing the Stroop task in the study by Clawson et al. (in press) suggested that on many trials subjects are aware of first thinking the word then naming the color. Speed-up on the Stroop task could therefore be attributed to faster retrieval and subvocalization of the words, allowing quicker onset of the color-name vocalization. This explanation poses no challenge to previous views of skilled performance as direct memory retrieval. The findings of Dulaney and Rogers (1994) counter this explanation, however. Dulaney and Rogers found that word-reading performance on incongruent stimuli previously used in Stroop practice was slower than it had been before practice, a finding incompatible with either faster memory retrieval or faster subvocalization (because speed of subvocalization has been found to be identical to speed of overt vocalization, Landauer, 1962)

Another area addressed by this study has particular implications for the Stroop effect. In Experiment 1, Stroop practice, but not simple color-naming practice, aided later Stroop performance, a finding that has interesting implications for theories explaining the Stroop effect. Both the horse-race model and automaticity models account for Stroop interference with the difference in speed of processing or automaticity between reading and color-naming. If the difference were reduced by hastening color-naming, then it would seem that the interference should be reduced. In Experiment 1 lines-trained subjects identified colors more quickly after training than did the control group but were no faster on the Stroop task. This could imply that neither the horse-race nor the automaticity theory can properly account for the Stroop effect. However, it is possible to

incorporate into those theories a threshold level of difference between color naming and word reading above which there is great interference and below which there is not. According to that assumption, color naming in the lines-trained subjects would have been fast but not fast enough to lower the difference below threshold.

A further question of interest is that of what are the relevant retrieval cues (or anti-retrieval cues). If there is specificity to colors, Experiment 3 makes it clear that the specificity is not to perceptual aspects of those colors. Based on Clawson et al.'s (in press) finding that orthographic manipulations did not affect skilled performance of the Stroop task by a small group of subjects, and based on Musen and Squire's (1993) finding that switching typecase did not affect skilled Stroop performance in their sample of amnesic and normal adults, it appears that the specificity to words is not dependent on word form. Reisberg et al.'s (1980) finding of no transfer of practice effects to homophonic words suggests further that word specificity is not due to phonemic aspects of the words. All of these differences are treated as superficial in the subjects' skilled performance. An area of further research is to determine what makes these changes superficial.

While the specificity to words found in this study challenges the theory that skilled performance depends on direct retrieval of correct responses from memory, it does support the basis of the theory. Specificity to the practiced words supports the role of memory for those words in skilled Stroop performance. Where previous embodiments of the theory allowed a role for increased memory retrieval, the theory can now include a role for decreased memory retrieval, as well. The addition of decreased memory retrieval to theories emphasizing the role of memory in skilled performance expands the domain of those theories, allowing better understanding of the processes underlying improvements in performance due to practice.

## Author Notes

This technical report originated as my dissertation. I wish to thank my committee chair, Alice Healy, and committee members Lyle Bourne, Jack Werner, Gary McClelland, and Clayton Lewis. Thanks also go to my undergraduate research assistants, Ross Artwohl, Alyson Kolber, Jodi Schaeffer. I am further grateful to Keizo Shinomori for measuring the luminance and the chromaticity of the colors used in Experiment 3.

# REFERENCES

Brown, W. (1915). Practice in associating color names with colors. Psychological Review, 22, 45-55.

Clawson, D. M., King, C. L., Healy, A. F., & Ericsson, K. A. (in press). Training and retention of the classic Stroop task: Specificity of practice effects. In A. F. Healy & L. E. Bourne, Jr. (Eds.), Learning and memory of knowledge and skills. Newbury Park, CA: Sage.

Connor, A., Franzen, M., & Sharp, B. (1988). Effects of practice and differential instructions on Stroop performance. The International Journal of Clinical Neuropsychology, 10, 1-4.

Dulaney, C. L., & Rogers, W. A. (1994). Mechanisms underlying reduction in Stroop interference with practice for young and old adults. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20, 470-484.

Dyer, F. N. (1973). The Stroop phenomenon and its use in the study of perceptual, cognitive, and response processes. Memory & Cognition, 1, 106-120.

Fendrich, D. W., Healy, A. F.., & Bourne, L. E.., Jr. (1993). Mental arithmetic: Training and retention of multiplication skill. In C. Izawa (Ed.), Cognitive psychology applied (pp. 111-133). Hillsdale, New Jersey: Erlbaum.

Flowers, J. H., & Stoup, C. M. (1977). Selective attention between words, shapes, and colors in speeded classification and vocalization tasks. Memory & Cognition, 5, 299-307

Harbeson, M. M., Krause, M., Kennedy, R. S., & Bittner, A. C. (1982). The Stroop as a performance evaluation test for environmental research. Journal of Psychology, 111, 223-233.

Healy, A. F., Fendrich, D. W., Crutcher, R. J., Wittman, W. T., Gesi, A. T., Ericsson, K. A., & Bourne, L. E., Jr. (1992) The long-term retention of skills. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), From learning processes to cognitive processes: Essays in honor of William K. Estes (Vol. 2, pp. 87-118). Hillsdale, NJ: Erlbaum.

Healy, A. F., King, C. L., Clawson, D. M., Sinclair, G. P., Rickard, T. C., Crutcher, R. J., Ericsson, K. A., & Bourne, L. E., Jr. (in press). Optimizing the long-term retention of skills. In A. F.

Healy & L. E. Bourne, Jr. (Eds.), Learning and memory of knowledge and skills. Newbury Park, CA: Sage.

Jensen, A. R., & Rohwer, W. D., Jr. (1966). The Stroop color-word test: A review. Acta Psychologica, 25, 36-93.

Landauer, T. K. (1962). Rate of implicit speech. Perceptual and Motor Skills, 15, 646.

Logan, G. D. & Klapp, S. T. (1991). Automatizing alphabet arithmetic: I. Is extended practice necessary to produce automaticity? Journal of Experimental Psychology: Learning, Memory and Cognition, 17, 179-195.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. Psychological Bulletin, 109, 163-203.

MacLeod, C. M. (1992). The Stroop task: The "gold standard" of attentional measures. Journal of Experimental Psychology: General, 121, 12-14.

MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. Journal of Experimental Psychology: Learning, Memory, and Cognition, 14, 126-135.

Masson, M. E. J. (1986). Identification of typographically transformed words: Instance-based skill acquisition. Journal of Experimental Psychology: Learning, Memory, and Cognition, 12, 479-488.

Melara, R. D., & Mounts, J. R. W. (1993). Selective attention to Stroop dimensions: Effects of baseline discriminability, response mode, and practice. Memory & Cognition, 21, 627-645.

Ménard-Buteau, C., & Cavanagh, P. (1984). Localization of the form/colour interference at the perceptual level in a Stroop task with stimuli drawings. Canadian Journal of Psychology, 38, 421-439.

Musen, G., & Squire, L. R. (1993). Implicit learning of color-word associations using a Stroop paradigm. Journal of Experimental Psychology: Learning, Memory, and Cognition, 19, 789-798.

National Research Council. (1981). Procedures for testing color vision: Report of working group 41. Washington, DC: National Academy Press.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), Cognitive skills and their acquisition (pp. 1-55). Hillsdale, NJ: Erlbaum

Proctor, J. D., & Healy, A. F. (in press). Acquisition and retention of skilled letter detection. In A. F. Healy & L. E. Bourne, Jr. (Eds.), Learning and memory of knowledge and skills. Newbury Park, CA: Sage.

Reisberg, D., Baron, J., & Kemler, D. G. (1980). Overcoming Stroop interference: The effects of practice on distractor potency. Journal of Experimental Psychology: Human Perception and Performance, 6, 140-150.

Rickard, T. C., Healy, A. F., & Bourne, L. E., Jr. (in press). On the cognitive structure of basic arithmetic skills: Operation, order, and symbol transfer effects. Journal of Experimental Psychology: Learning, Memory, and Cognition.

Rickard, T. C., & Bourne, L. E., Jr. (in press). An identical elements model of basic arithmetic skills. In A. F. Healy & L. E. Bourne, Jr. (Eds.), Learning and memory of knowledge and skills. Newbury Park, CA: Sage.

Roe, W. T., Wilsoncroft, W. E., & Griffiths, R. S. (1980). Effects of motor and verbal practice on the Stroop task. Perceptual and Motor Skills, 50, 647-650.

Sacks, T. L., Clark, C. L., Pols, R. G., & Geffen, L. B. (1991). Comparability and stability of performance of six alternate forms of the Dodrill-Stroop colour-word test. The Clinical Neuropsychologist, 5, 220-225.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. Psychological Science, 3, 207-217.

Schneider, V. I., Healy, A. F., & Gesi, A. T. (1991). The role of phonetic processes in letter detection: A reevaluation. Journal of Memory and Language, 30, 294-318.

Schneider, W. (1988). Micro Experimental Laboratory: An integrated system for IBM PC compatibles. Behavior Research Methods, Instruments, & Computers, 20, 206-217.

Shor, R. E., Hatch, R. P., Hudson, L. J., Landrigan, D. T., & Shaffer, H. J. (1972). Effect of practice on a Stroop-like spatial directions task. Journal of Experimental Psychology, 94, 168-172.

Simon, J. R. (1990) The effects of an irrelevant directional cue on human information processing. In R. W. Proctor & T. G. Reeve (Eds.), Stimulus-response compatability: An integrated perspective (pp. 31-86). Amsterdam, Netherlands: North-Holland.

Singley, M. K., & Anderson, J. R. (1989). The transfer of cognitive skill. Cambridge, MA: Harvard University Press.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. Journal of Experimental Psychology, 18, 643-662.

Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions (I). Psychological Review, 8, 247-261.

White, L. C. (1978). Interference proneness and the ability to shift attention in old age (Doctoral dissertation, University of Notre Dame, 1978). Dissertation Abstracts International, 39, 2549-B.

Whittlesea, B. W. A., & Brooks, L. R. (1988). Critical influence of particular experiences in the perception of letters, words, and phrases. Memory & Cognition, 16, 387-399.

Wyszecki, G., & Stiles, W. S. (1982). Color science: Concepts and methods, quantitative data and formulae (2nd Edition). New York: John Wiley & Sons.

# APPENDIX A
# EXPERIMENT 1 PATTERN OF REACTION TIMES DURING PRACTICE

The pattern of reaction times during practice for both the lines group and the Stroop group were consistent with power law speed-up. The data, however, were noisy enough that other functions could not be ruled out as equally well fitted.

Examining the pattern of improvement in reaction times due to practice was carried out in two ways, measuring estimates of fit ($r^2$) and viewing the data for systematic deviations from the proposed functions. The three functions fitted to the practice data were a linear function, a simple power function, and a simple exponential function. Just as linear functions appear as straight lines when plotted in normal Cartesian coordinates (x, y), Newell and Rosenbloom (1981) pointed out that power functions appear as straight lines in log-log coordinates (log x, log y) and exponential functions appear as straight lines in semilog space (x, log y). Therefore, each best-fitting function and the data were viewed in the appropriate space (nonlog, log-log, or semilog) to look for systematic deviations from linearity.

Mean log reaction times for each of the 40 practice blocks across subjects were calculated separately for the lines and Stroop groups; for comparison to the linear function, these mean log reaction times were converted into ms. Thus, for each group there were 40 data points, each corresponding to the mean of the eight subjects' individual mean log reaction times on the 12 trials of a given practice block. Equations for the simple power function, the simple exponential function, and a linear function were then fitted to the data for each group. No attempt was made to distinguish between the power function and its special case, the hyperbolic function, because of the great noise in the data. The simple power function and simple exponential function, in terms of log reaction time, and the linear function, in terms of ms reaction time, were as follows:

$$\log RT = \log a - b\,(\log \text{block})$$
$$\log RT = \log c - d\,(\text{block})$$
$$RT = f\,(\text{block}) + g$$

Examining the data for systematic deviations from the functions was accomplished by first viewing the data in log-log (i.e., logRT-log block) coordinates, in which power functions appear as straight lines, then viewing the data in semilog (i.e., logRT-block) coordinates, in which exponential functions appear as straight lines, and finally viewing the data in nonlog (i.e. RT-block) coordinates, in which linear functions appear as straight lines.

For the lines training group, Table A-1 shows the best fitting power, exponential, and linear functions for the practice reaction times. As can be seen in Figures A-1, A-2, and A-3, the reason for the relatively low and indistinguishable estimates of fit ($r^2$) is noise in the data. No systematic deviations from linearity were apparent in

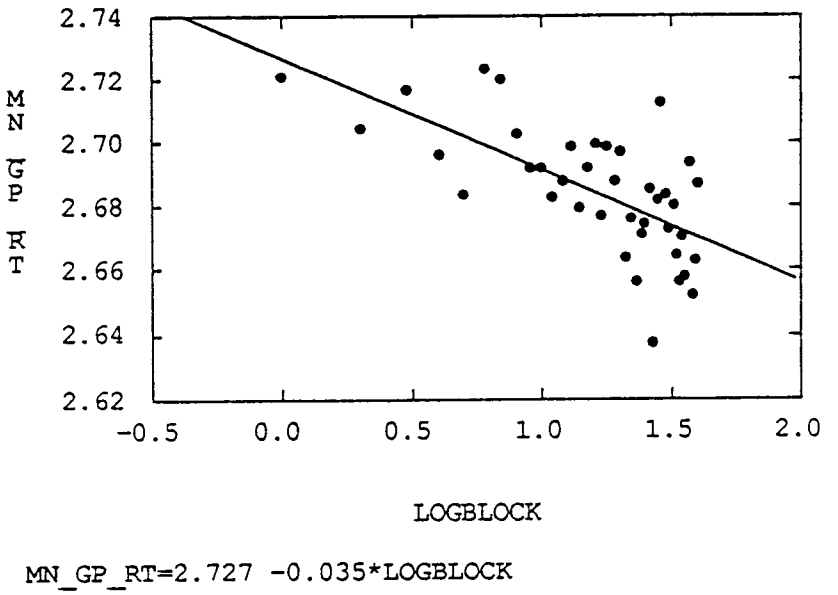



Figure A-1. Experiment 1 mean log reaction times for the lines-trained group (mn_gp_rt) for each log block in practice and the best-fitting power function for those data. Note that power functions appear linear in log-log space.

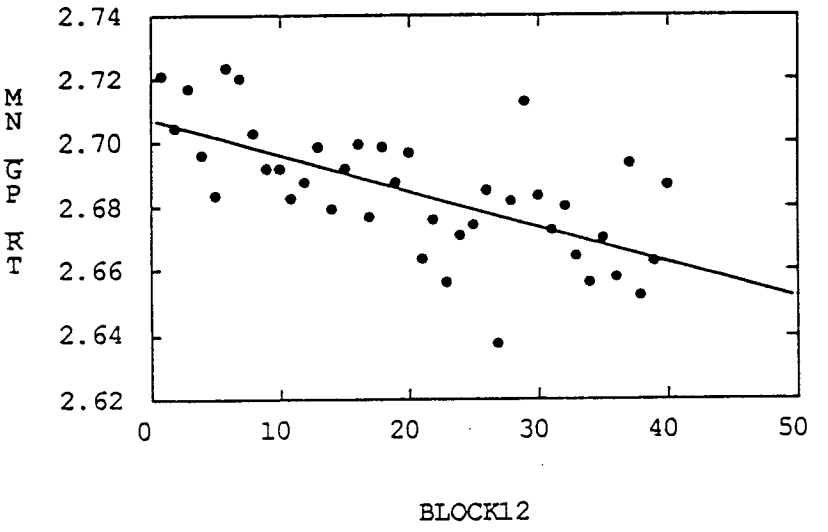



Figure A-2. Experiment 1 mean log reaction times for the lines-trained group (mn_gp_rt) for each block (block12) in practice and the best-fitting exponential function for those data. Note that exponential functions appear linear in semilog space.

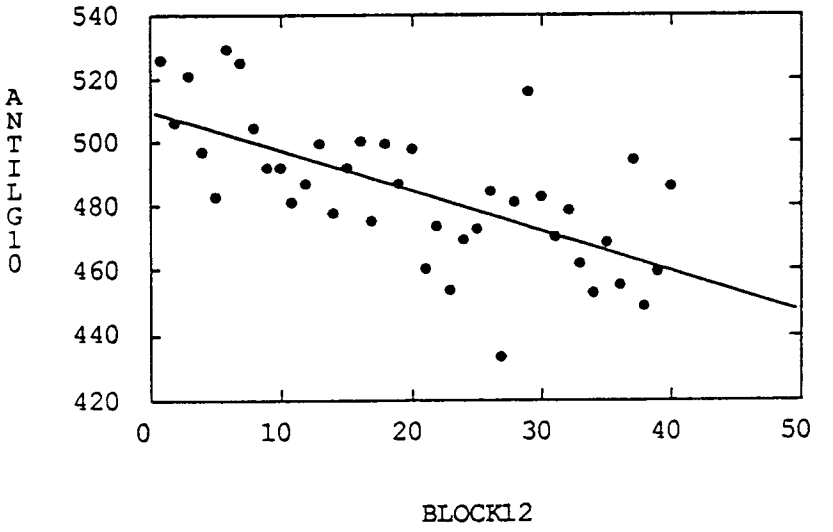



Figure A-3. Experiment 1 mean reaction times in ms (antilg10) for the lines-trained group for each block (block12) in practice and the best-fitting linear function for those data.

any of the spaces, but the noise is so great as to make any systematic deviations difficult to ascertain. The noise was large in each individual subject's data as well, as can be seen in the graphs at the end of this Appendix. Note, however, that all lines subjects did show some tendency toward decreasing reaction times with practice.

Table A-1
Simple power, simple exponential, and linear functions that best fit practice reaction times across lines-trained subjects, their estimates of fit to the data ($r^2$), and the space in which the functions approximate straight lines.

| **Function** | **Space** | **Best-fitting function** | $r^2$ |
|---|---|---|---|
| Simple power | log-log | logRT=2.727 -.035(log block) | .44 |
| Simple exponential | semilog | logRT=2.707 - .001 (block) | .43 |
| Linear | nonlog | RT=510 - 1.246 (block) | .43 |

Noise was also a noticeable feature of practice reaction times for the Stroop-trained group. The noise is apparent in each individual subject's data as well, as can be seen in the individual subject graphs which follow. For the Stroop training group, Table A-2 shows the best fitting power, exponential, and linear functions for the practice reaction times; the power function has the best fit. As can be seen in Figures A-4, A-5, and A-6, the reason for the relatively low estimates of fit ($r^2$) is noise in the data. Note, however, that all subjects except Subject 16 show decreasing reaction times with practice.

Table A-2
Simple power, simple exponential, and linear functions that best fit practice reaction times across Stroop-trained subjects, their estimates of fit to the data ($r^2$), and the space in which the functions approximate straight lines.

| **Function** | **Space** | **Best-fitting function** | $r^2$ |
|---|---|---|---|
| Simple power | log-log | logRT=2.851 -.066(log block) | .68 |
| Simple exponential | semilog | logRT=.810 - .002 (block) | .50 |
| Linear | nonlog | RT=646 -2.599 (block) | .50 |

The following graphs display each individual Experiment 1 subject's mean reaction times (in ms) for each block of practice. Graphs for lines-trained subjects (training condition=1) are presented

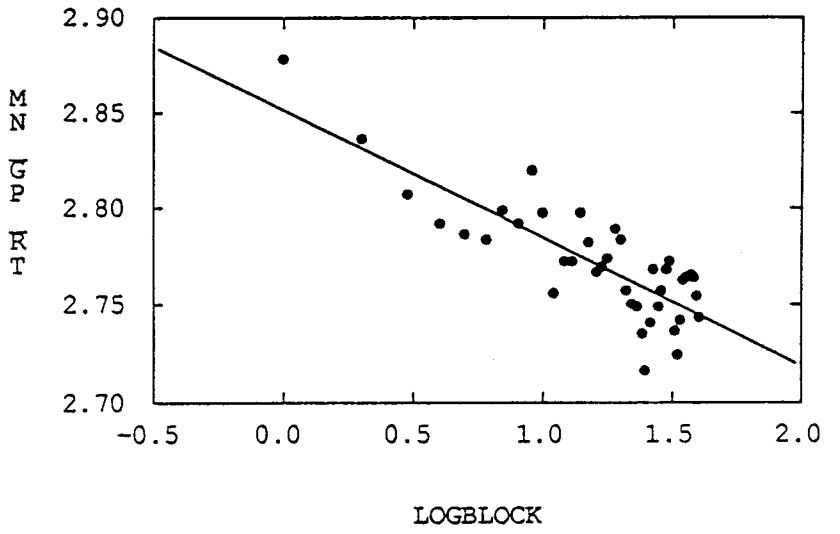



Figure A-4. Experiment 1 mean log reaction times for the Stroop-trained group (mn_gp_rt) for each log block in practice and the best-fitting power function for those data. Note that power functions appear linear in log-log space.

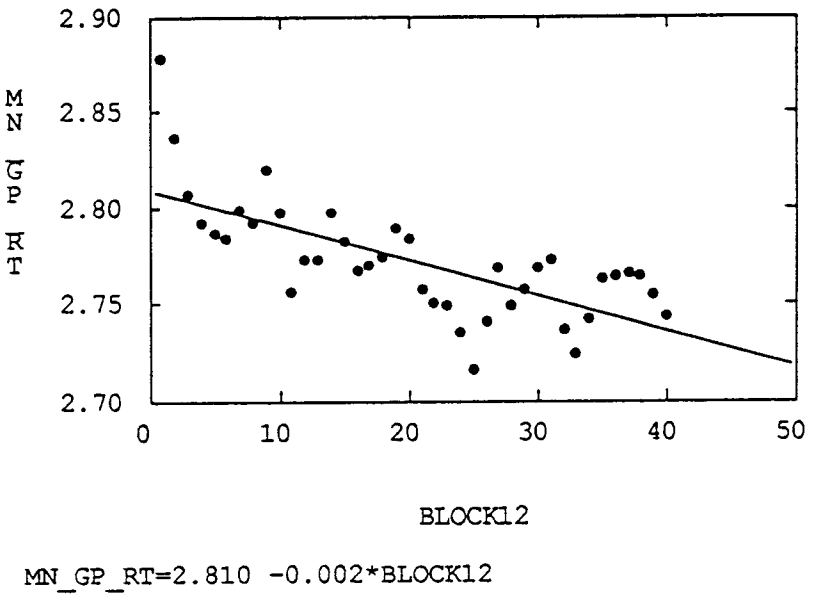



Figure A-5. Experiment 1 mean log reaction times for the Stroop-trained group (mn_gp_rt) for each block (block12) in practice and the best-fitting exponential function for those data. Note that exponential functions appear linear in log-log space.

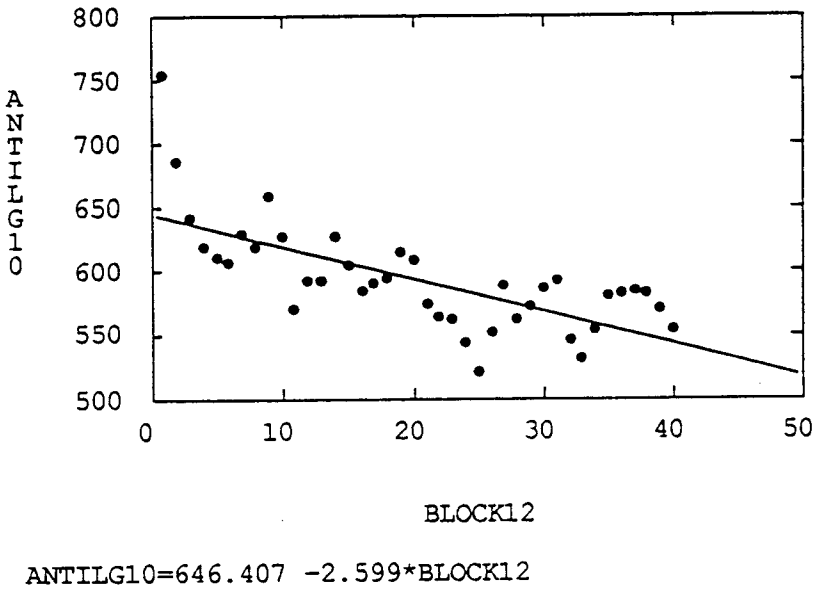
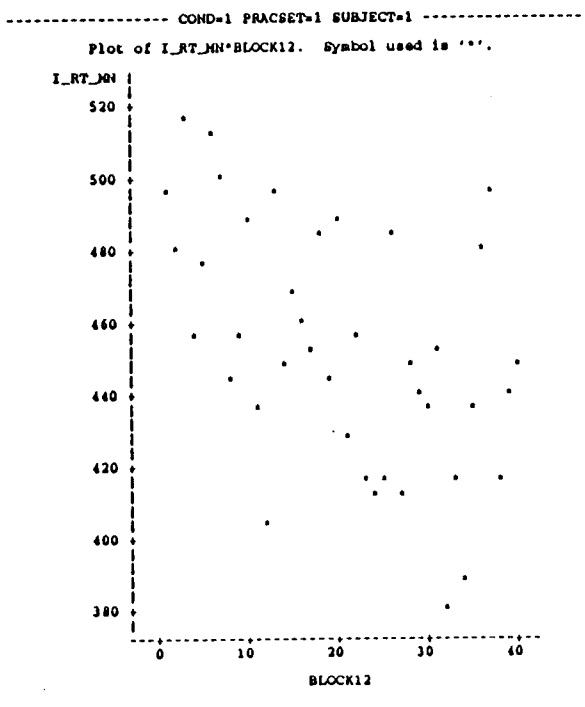
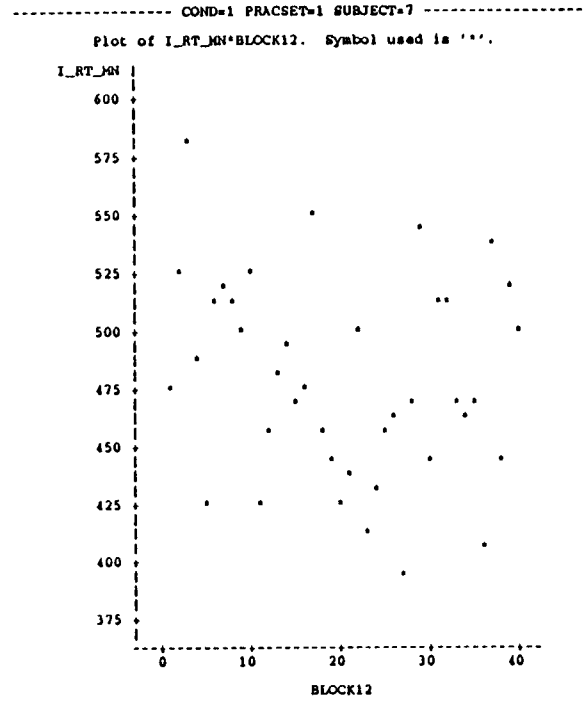
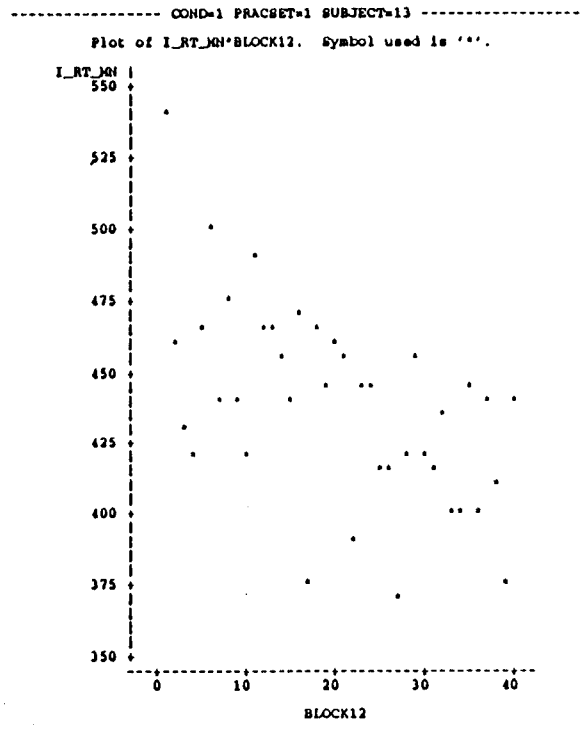


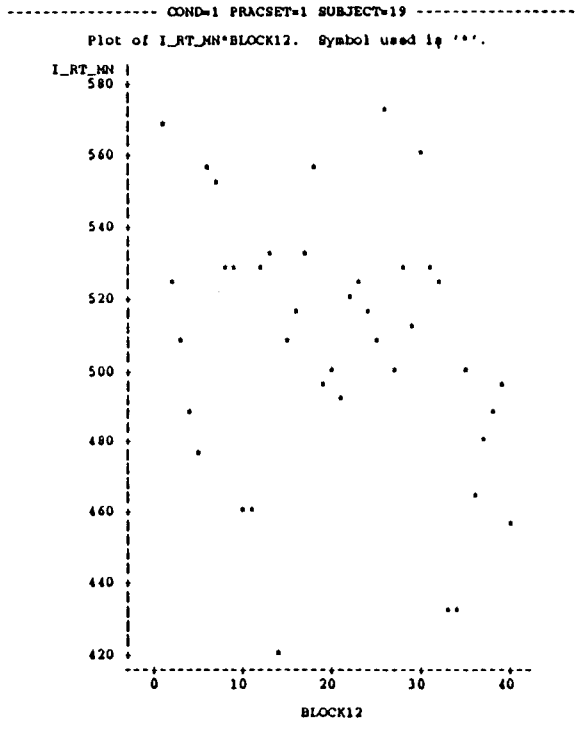

Figure A-6. Experiment 1 mean reaction times in ms (antilg10) for the Stroop-trained group for each block (block12) in practice and the best-fitting linear function for those data.

first, followed by graphs for Stroop-trained subjects (training condition=2). Within each training condition, graphs for subjects practicing on the pink-orange-blue set (pracset=1) precede graphs for subjects practicing on the red-green-purple set (pracset=2).


---------- COND=1 PRACSET=1 SUBJECT=1 ----------
Plot of I_RT_MN*BLOCK12. Symbol used is '*'.
I_RT_MN
520
500
480
460
440
420
400
380
0
10
20
30
40
BLOCK12

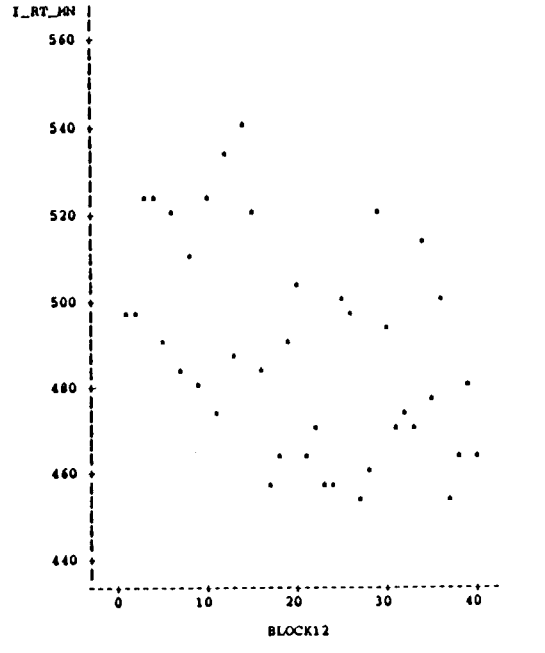

---------- COND=1 PRACSET=1 SUBJECT=7 ----------
Plot of I_RT_MN*BLOCK12. Symbol used is '*'.
I_RT_MN
600
575
550
525
500
475
450
425
400
375
0
10
20
30
40
BLOCK12


---------- COND=1 PRACSET=1 SUBJECT=13 ----------
Plot of I_RT_MN*BLOCK12. Symbol used is '*'.
I_RT_MN
550
525
500
475
450
425
400
375
350
0
10
20
30
40
BLOCK12


---------- COND=1 PRACSET=1 SUBJECT=19 ----------
Plot of I_RT_MN*BLOCK12. Symbol used is '*'.
I_RT_MN
580
560
540
520
500
480
460
440
420
0
10
20
30
40
BLOCK12

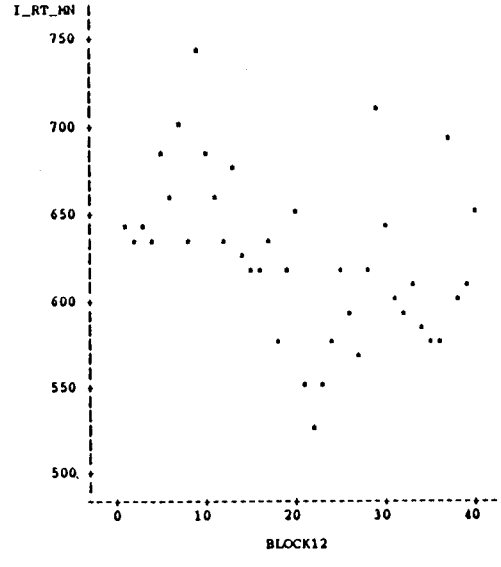
---------- COND=1 PRACSET=2 SUBJECT=2 ----------

Plot of I_RT_MN*BLOCK12. Symbol used is '*'.



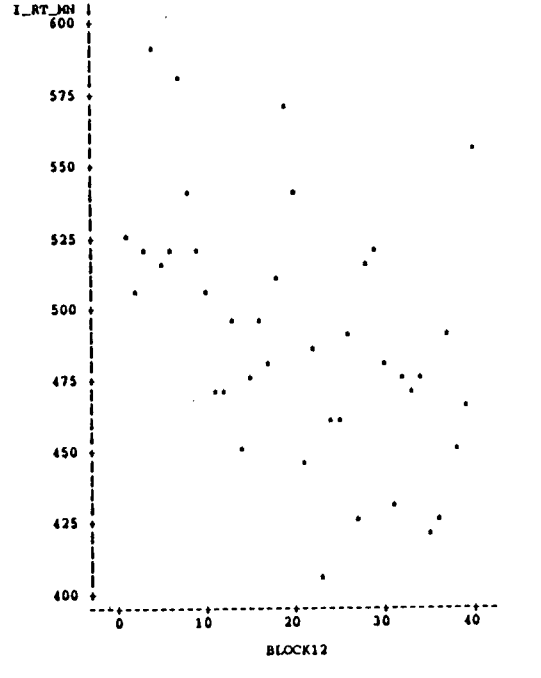

---------- COND=1 PRACSET=2 SUBJECT=8 ----------

Plot of I_RT_MN*BLOCK12. Symbol used is '*'.



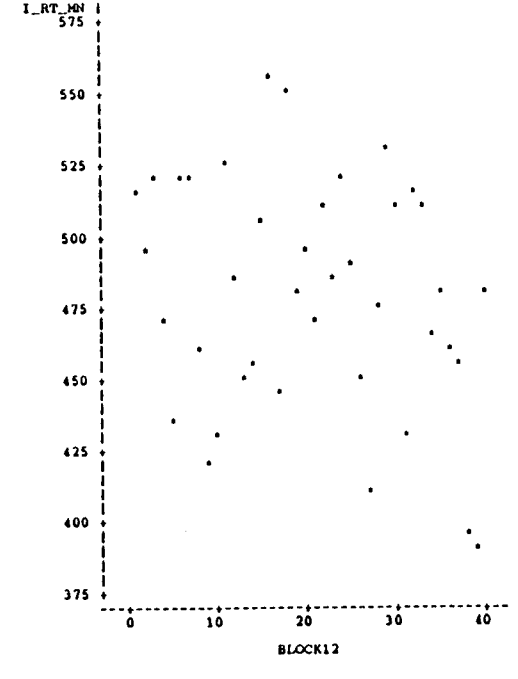
---------- COND=1 PRACSET=2 SUBJECT=14 ----------

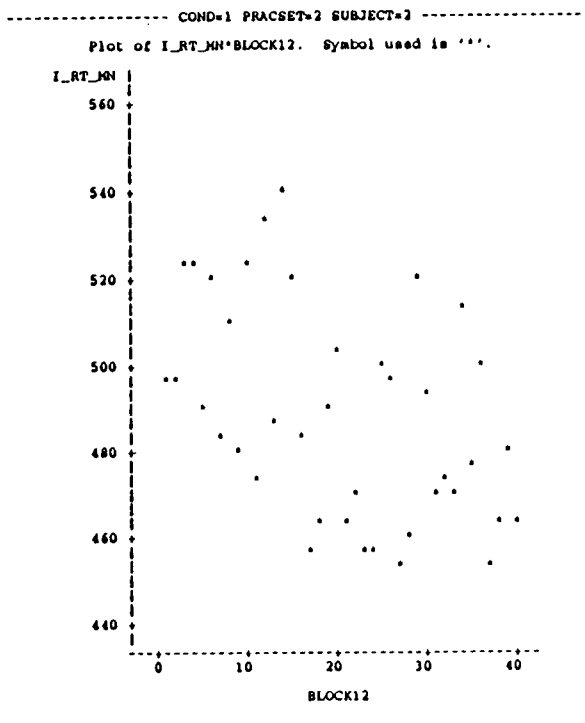
Plot of I_RT_MN*BLOCK12. Symbol used is '*'.

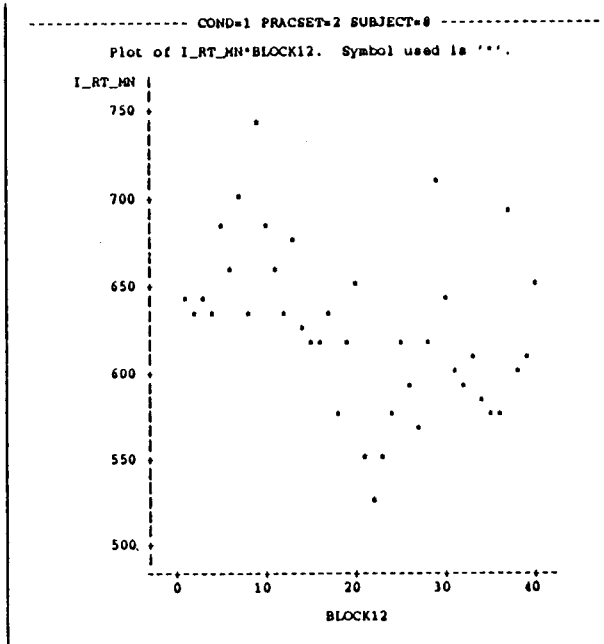


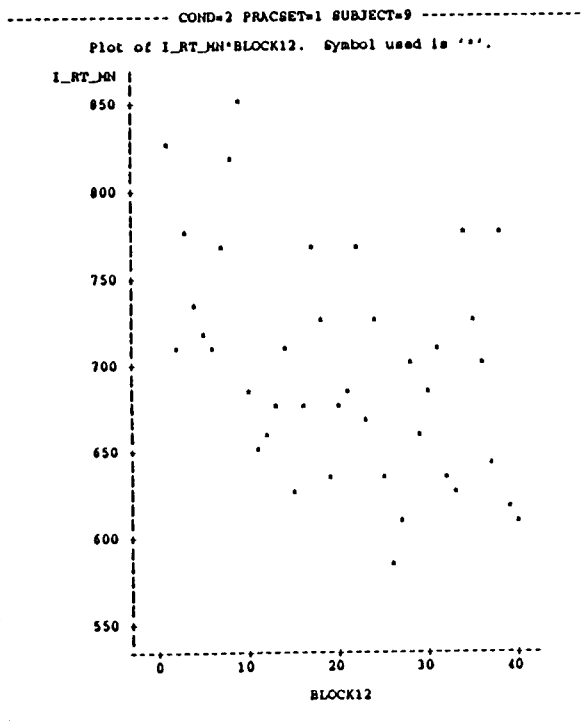

---------- COND=1 PRACSET=2 SUBJECT=20 ----------

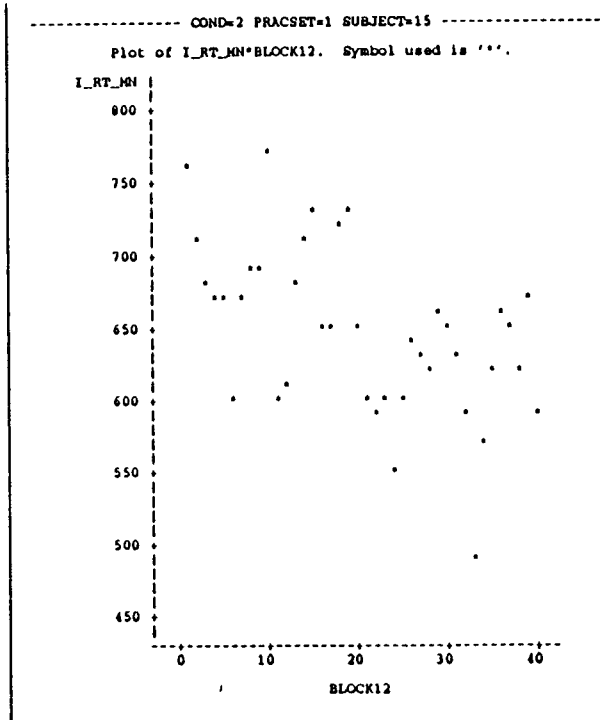
Plot of I_RT_MN*BLOCK12. Symbol used is '*'.

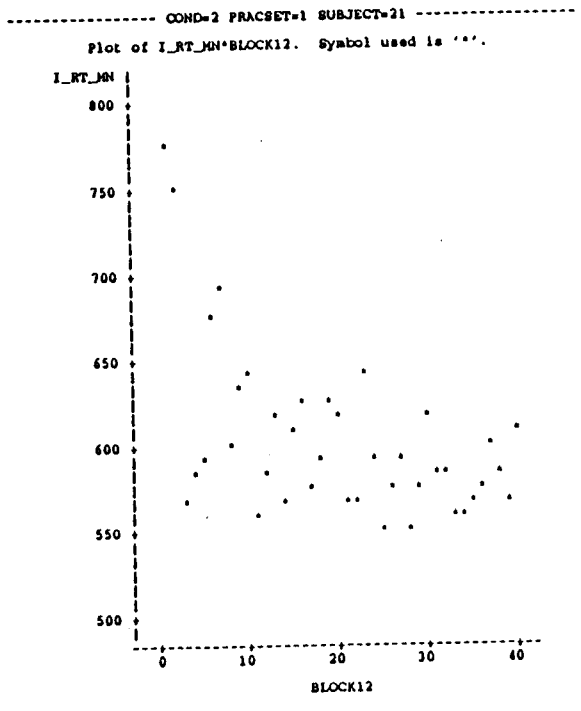


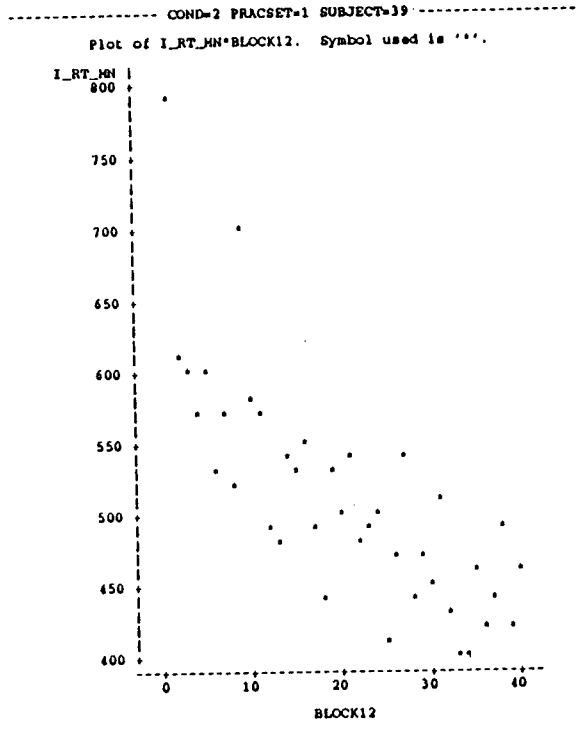

---------- COND=1 PRACSET=2 SUBJECT=2 ----------
Plot of I_RT_MN*BLOCK12. Symbol used is '*'.
I_RT_MN
560
540
520
500
480
460
440
0
10
20
30
40
BLOCK12

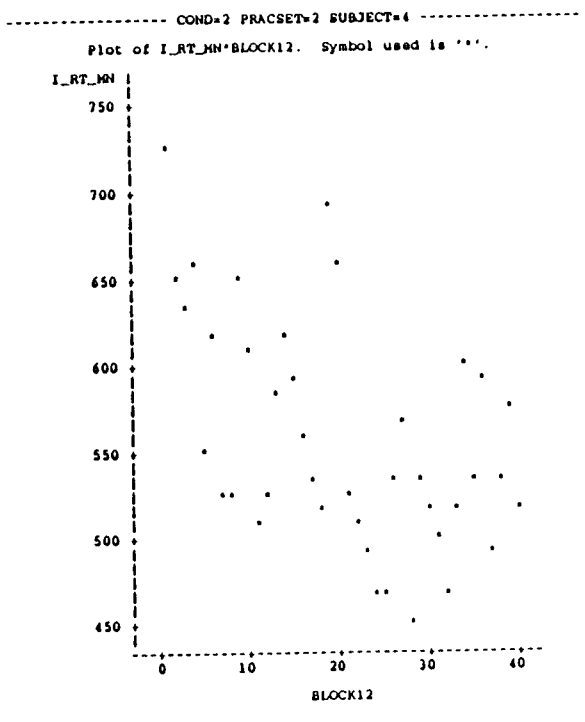

---------- COND=1 PRACSET=2 SUBJECT=8 ----------
Plot of I_RT_MN*BLOCK12. Symbol used is '*'.
I_RT_MN
750
700
650
600
550
500
0
10
20
30
40
BLOCK12

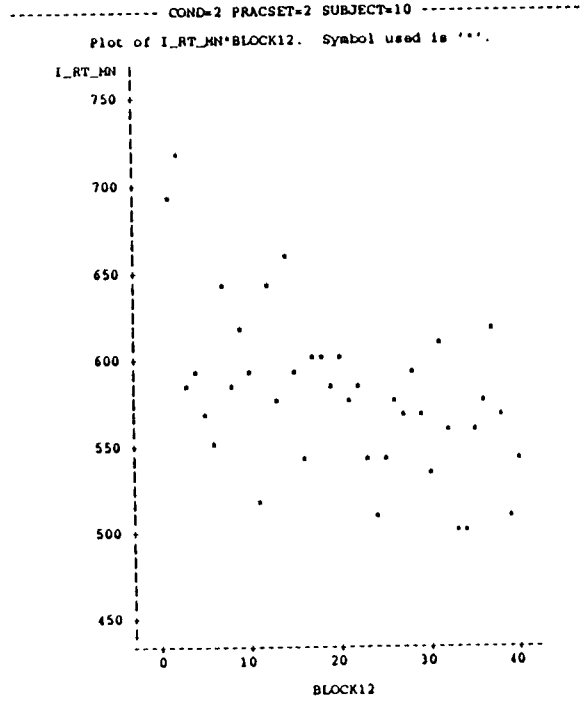

COND=2 PRACSET=1 SUBJECT=9
Plot of I_RT_MN*BLOCK12. Symbol used is '*'.
I_RT_MN
850
800
750
700
650
600
550
0
10
20
30
40
BLOCK12

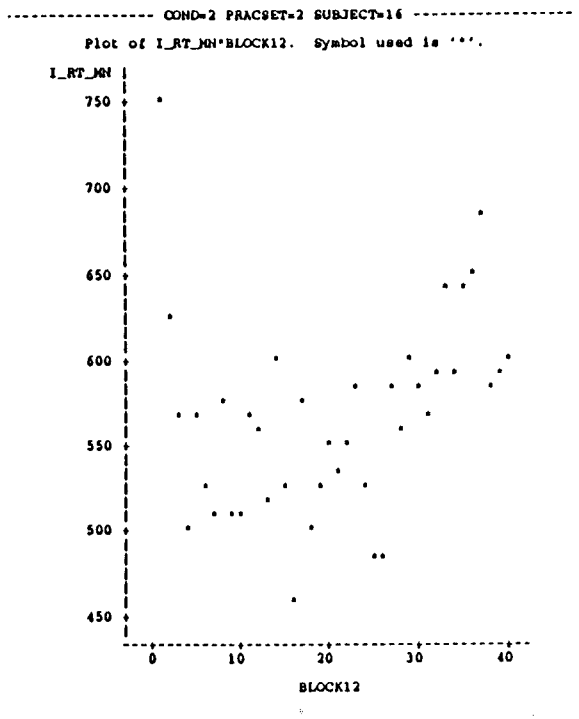

COND=2 PRACSET=1 SUBJECT=15
Plot of I_RT_MN*BLOCK12. Symbol used is '*'.
I_RT_MN
800
750
700
650
600
550
500
450
0
10
20
30
40
BLOCK12

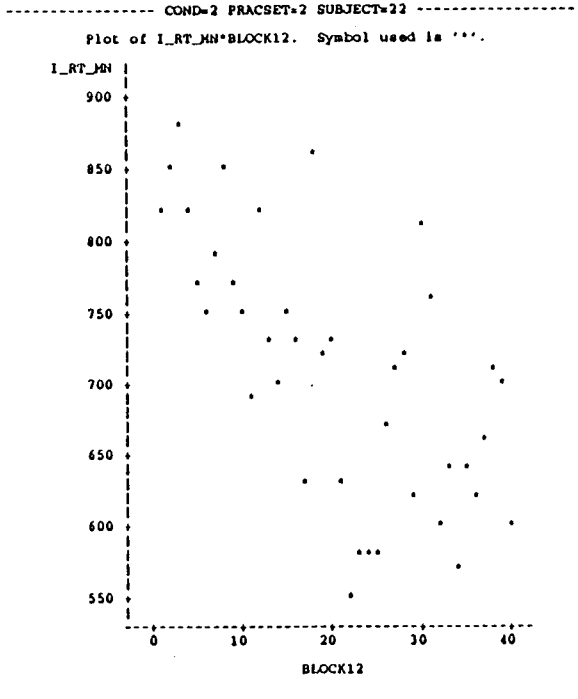

COND=2 PRACSET=1 SUBJECT=21
Plot of I_RT_MN*BLOCK12. Symbol used is '*'.
I_RT_MN
800
750
700
650
600
550
500
0
10
20
30
40
BLOCK12


COND=2 PRACSET=1 SUBJECT=39
Plot of I_RT_MN*BLOCK12. Symbol used is '*'.
I_RT_MN
800
750
700
650
600
550
500
450
400
0
10
20
30
40
BLOCK12


COND=2 PRACSET=2 SUBJECT=4
Plot of I_RT_MN*BLOCK12. Symbol used is '*'.
I_RT_MN
750
700
650
600
550
500
450
0
10
20
30
40
BLOCK12


COND=2 PRACSET=2 SUBJECT=10
Plot of I_RT_MN*BLOCK12. Symbol used is '*'.
I_RT_MN
750
700
650
600
550
500
450
0
10
20
30
40
BLOCK12


---------- COND=2 PRACSET=2 SUBJECT=16 ----------
Plot of I_RT_MN*BLOCK12. Symbol used is '*'.
I_RT_MN
750
700
650
600
550
500
450
0
10
20
30
40
BLOCK12


---------- COND=2 PRACSET=2 SUBJECT=22 ----------
Plot of I_RT_MN*BLOCK12. Symbol used is '*'.
I_RT_MN
900
850
800
750
700
650
600
550
0
10
20
30
40
BLOCK12

# APPENDIX B
# EXPERIMENT 2 PATTERN OF REACTION TIMES DURING PRACTICE

The pattern of reaction times during practice was again consistent with power law speed-up. Reduced noise in the data allowed some distinction both in terms of estimates of fit and in terms of systematic deviations from the functions, suggesting that the power law accounted for the pattern of practice speed-up better than did simple exponential or linear functions.

As in Experiment 1 the examination of the pattern of improvement in reaction times due to practice was carried out both by measuring estimates of fit ($r^2$) and by viewing the data for systematic deviations from straight lines in log-log, semi-log, and nonlog space. Mean log reaction times were calculated for each of the 40 practice blocks across subjects; thus, there are 40 data points, each corresponding to the mean of 16 subjects' individual mean log reaction times on the 12 trials of a given practice block. Equations for the simple power function, the simple exponential function, and a linear function were then fitted to the data for each group. No attempt was made to distinguish between the power function and its special case, the hyperbolic function.

Table B-1 displays the simple power, exponential, and linear functions that best fit the practice reaction times along with their correlations with the data. The power function yielded a substantially larger estimate of fit ($r^2$) than did the other functions, suggesting that the power function best describes the pattern of speed-up during Stroop practice.

Table B-1
Simple power, simple exponential, and linear functions that best fit practice reaction times across subjects, their estimates of fit to the data ($r^2$), and the space in which the functions approximate straight lines.

| **Function** | **Space** | **Best-fitting function** | $r^2$ |
|---|---|---|---|
| Simple power | log-log | logRT=2.88 -.075l (log block) | .77 |
| Simple exponential | semilog | logRT=2.88 - .002 (block) | .63 |
| Linear | nonlog | RT=687 - 3.24 (block) | .62 |

The estimates of fit ($r^2$) are higher in this experiment than they were for Stroop practice in Experiment 2 mainly because of less noise, due to aggregation over twice as many subjects (because only 8 of the Experiment 1 subjects practiced the Stroop task whereas all 16 of the

subjects in this experiment did so). The reduced noise also allows finer examination of the best-fitting functions in their appropriate spaces. Figure B-1 shows the data and best-fitting power function, in log-log coordinates in which power functions approximate a straight line. There are no systematic deviations from the straight line. In Figure B-2, showing the data and best-fitting exponential function in semilog coordinates, there is slower than predicted performance at initial blocks which indicates that the power function is more appropriate. Figure B-3, showing the data and best-fitting linear function, reveals systematic slower than predicted performance at both initial and later blocks, again indicating that a power function is more appropriate.

As in Experiment 1, subjects' individual patterns of reaction time during practice were noisy, as can be seen in the following graphs. Note, however, that all subjects (excepting subjects 9 and 16) show decreasing reaction times with practice.

The following graphs display each individual Experiment 2 subject's mean reaction times (in ms) for each block of practice. (Every subject in Experiment 2 was Stroop-trained, training condition=2.) Graphs for subjects practicing on the pink-orange-blue set (pracset=1) precede graphs for subjects practicing on the red-green-purple set (pracset=2).

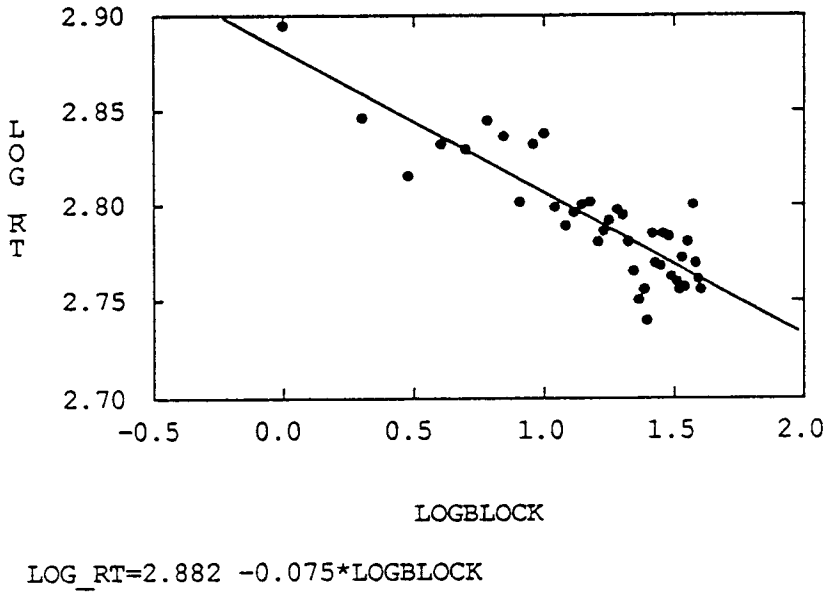



Figure B-1. Experiment 2 mean log reaction times (log_rt) for each log block in Stroop practice and the best-fitting power function for those data. Note that power functions appear linear in log-log space.

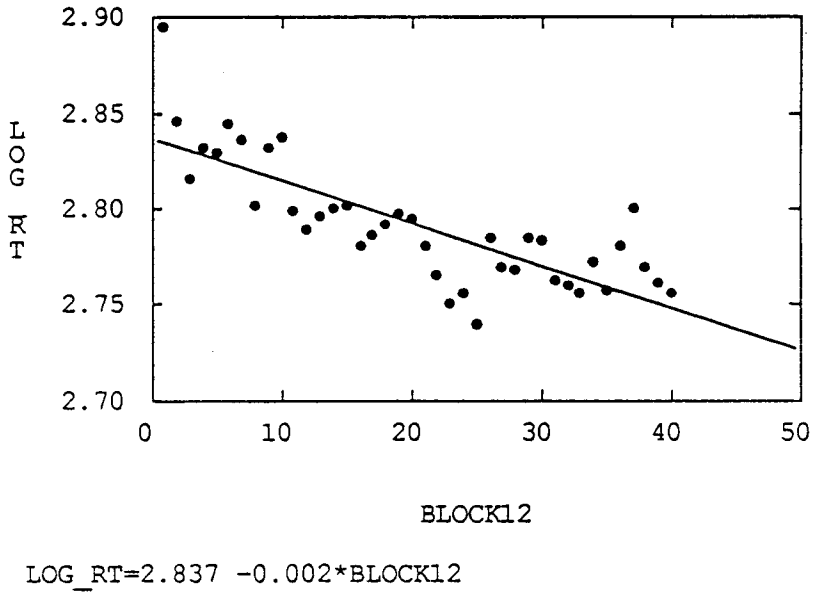



Figure B-2. Experiment 2 mean log reaction times (log_rt) for each block (block12) in Stroop practice and the best-fitting exponential function for those data. Note that exponential functions appear linear in log-log space.

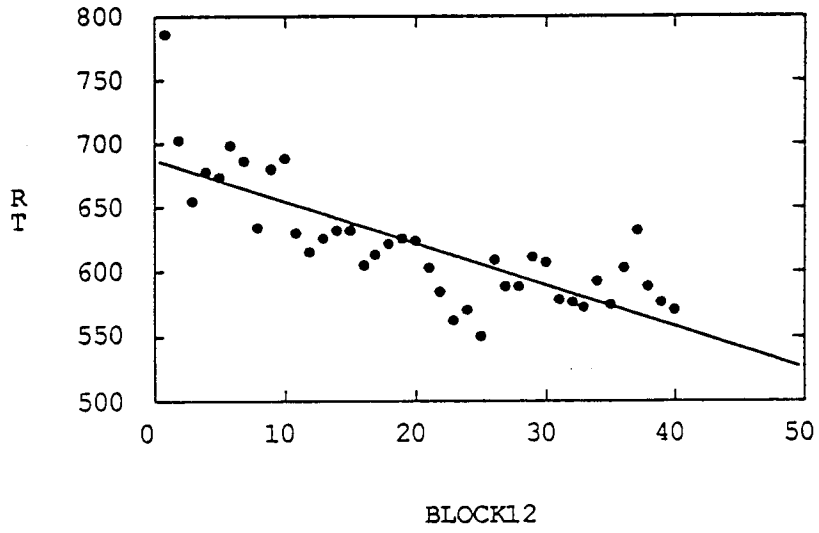
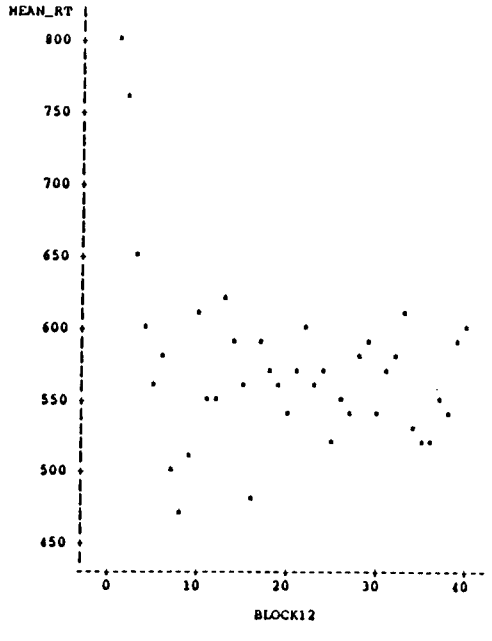



Figure B-3. Experiment 2 mean reaction times in ms (RT) for each block (block12) in Stroop practice and the best-fitting linear function for those data.

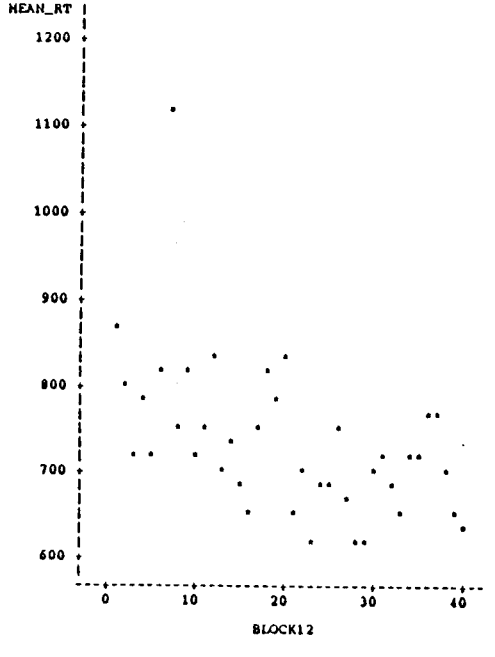
---------- COND=2 PRACSET=1 SUBJECT=1 ----------

Plot of MEAN_RT*BLOCK12. Symbol used is '*'.



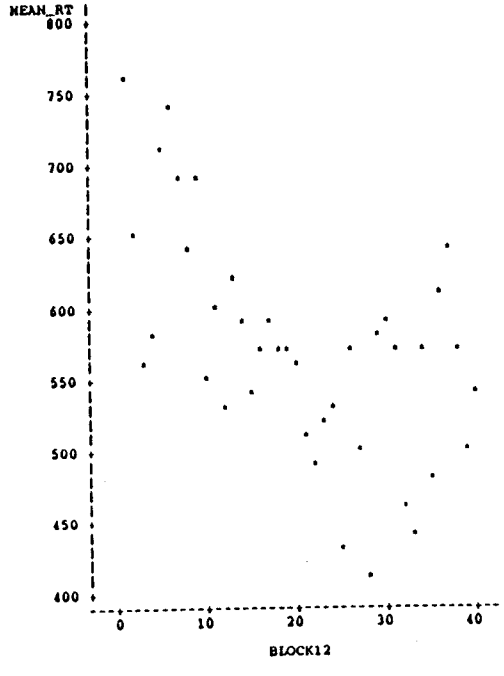

---------- COND=2 PRACSET=1 SUBJECT=2 ----------

Plot of MEAN_RT*BLOCK12. Symbol used is '*'.



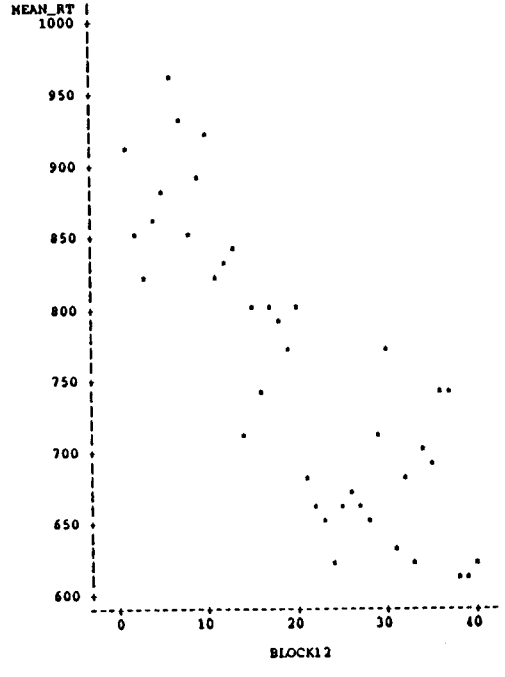
---------- COND=2 PRACSET=1 SUBJECT=10 ----------

Plot of MEAN_RT*BLOCK12. Symbol used is '*'.



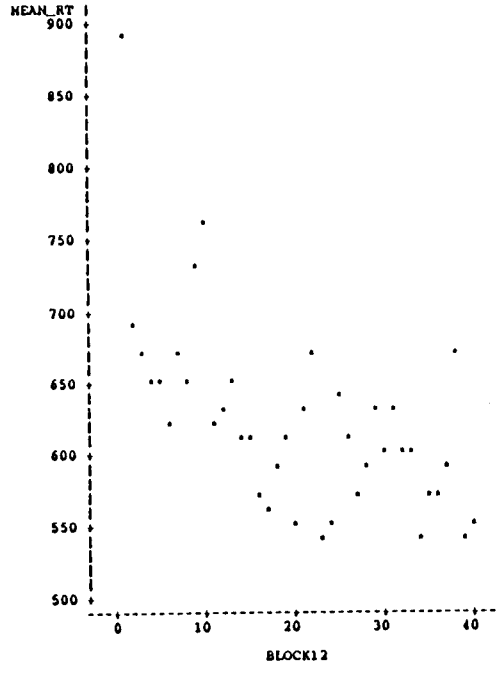

---------- COND=2 PRACSET=1 SUBJECT=11 ----------

Plot of MEAN_RT*BLOCK12. Symbol used is '*'.



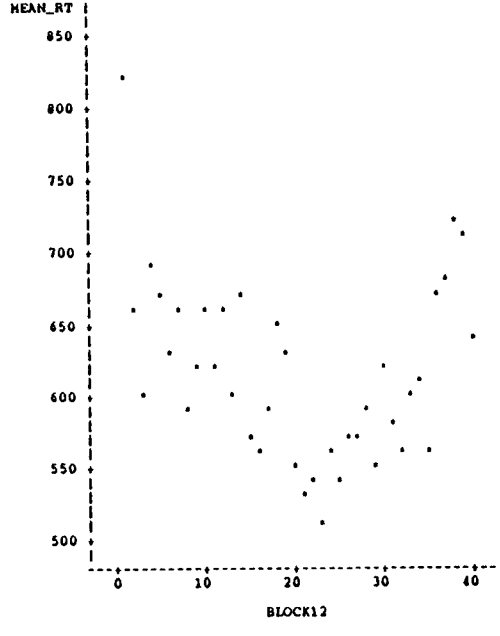
---------- COND=2 PRACSET=1 SUBJECT=4 ----------

Plot of MEAN_RT*BLOCK12. Symbol used is '*'.



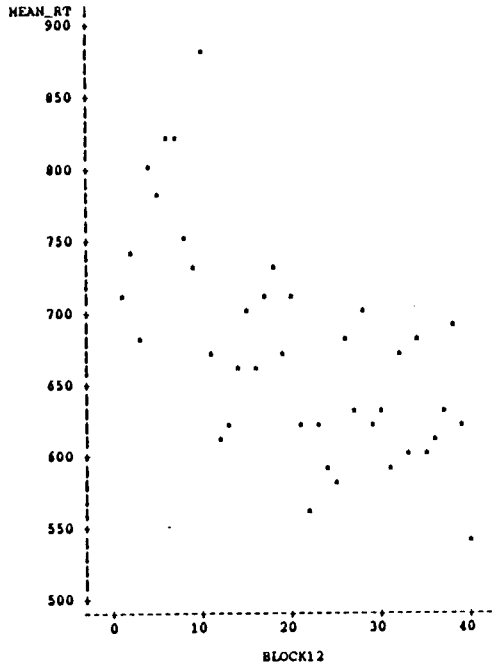

---------- COND=2 PRACSET=1 SUBJECT=9 ----------

Plot of MEAN_RT*BLOCK12. Symbol used is '*'.



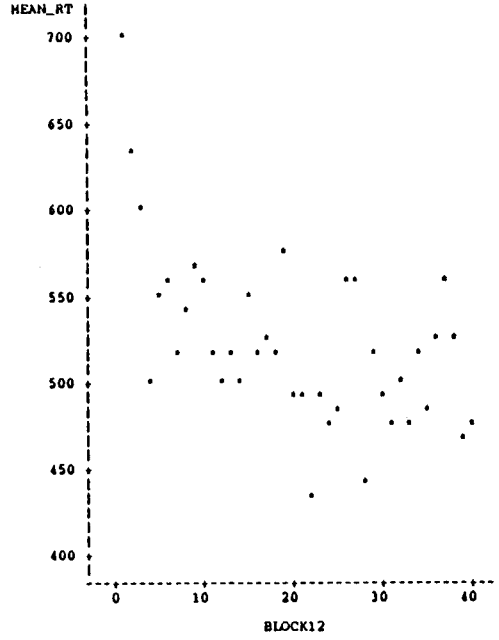
---------- COND=2 PRACSET=1 SUBJECT=12 ----------

Plot of MEAN_RT*BLOCK12. Symbol used is '*'.



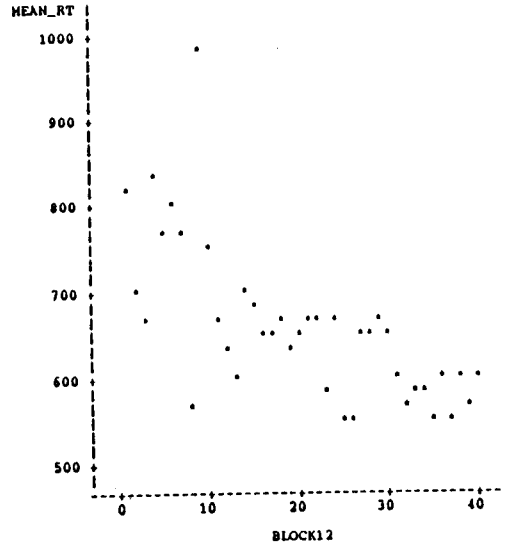

---------- COND=2 PRACSET=1 SUBJECT=19 ----------

Plot of MEAN_RT*BLOCK12. Symbol used is '*'.



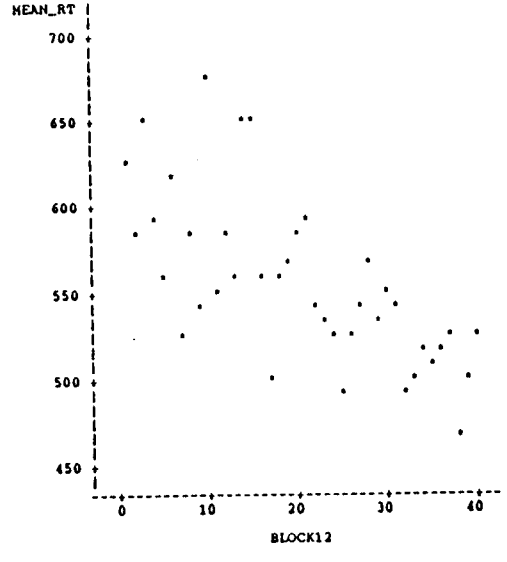
---------- COND=2 PRACSET=2 SUBJECT=5 ----------

Plot of MEAN_RT*BLOCK12. Symbol used is '*'.



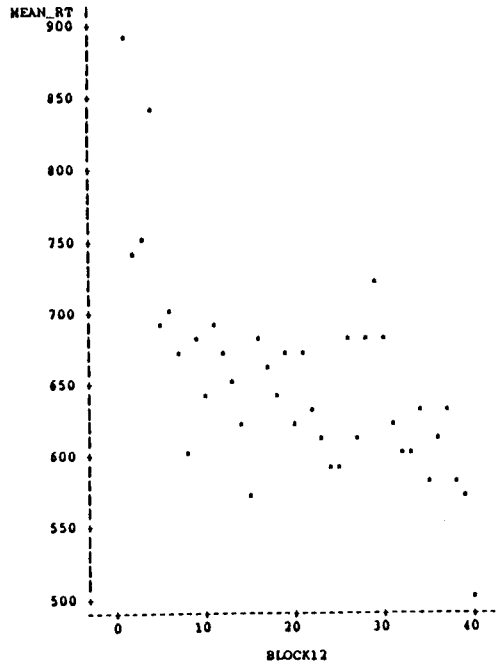

---------- COND=2 PRACSET=2 SUBJECT=6 ----------

Plot of MEAN_RT*BLOCK12. Symbol used is '*'.



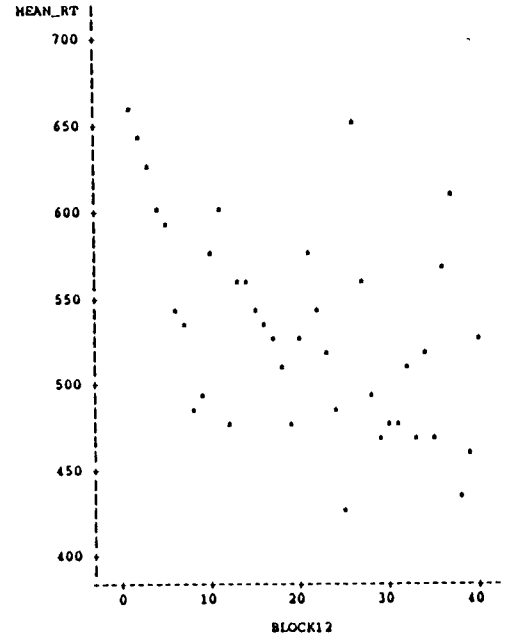
---------- COND=2 PRACSET=2 SUBJECT=7 ----------

Plot of MEAN_RT*BLOCK12. Symbol used is '*'.



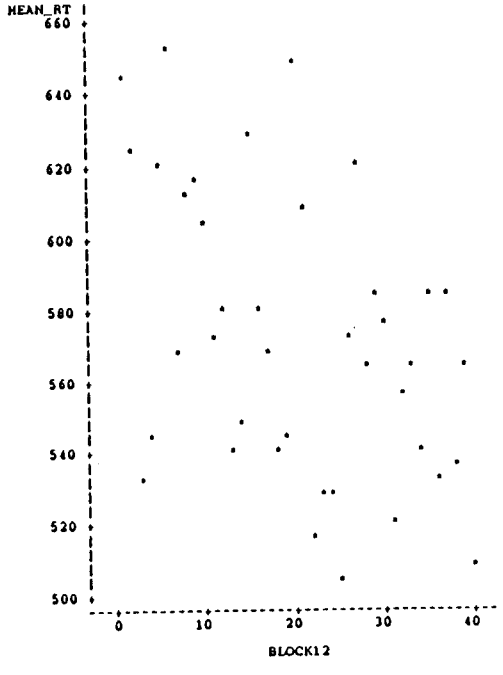

---------- COND=2 PRACSET=2 SUBJECT=8 ----------

Plot of MEAN_RT*BLOCK12. Symbol used is '*'.



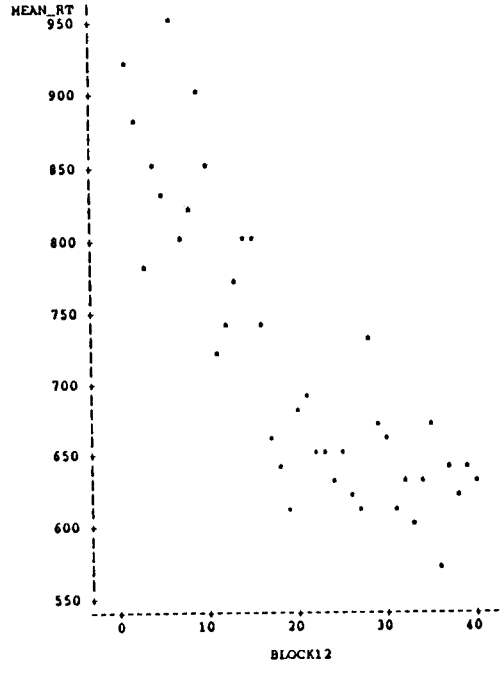
---------- COND=2 PRACSET=2 SUBJECT=13 ----------

Plot of MEAN_RT*BLOCK12. Symbol used is '*'.



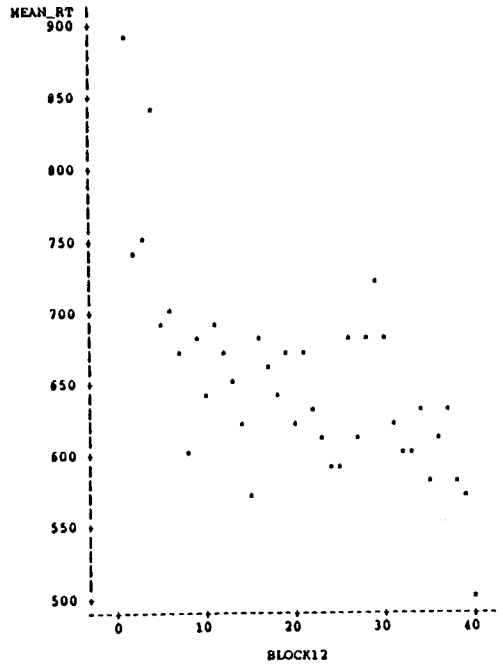

---------- COND=2 PRACSET=2 SUBJECT=14 ----------

Plot of MEAN_RT*BLOCK12. Symbol used is '*'.



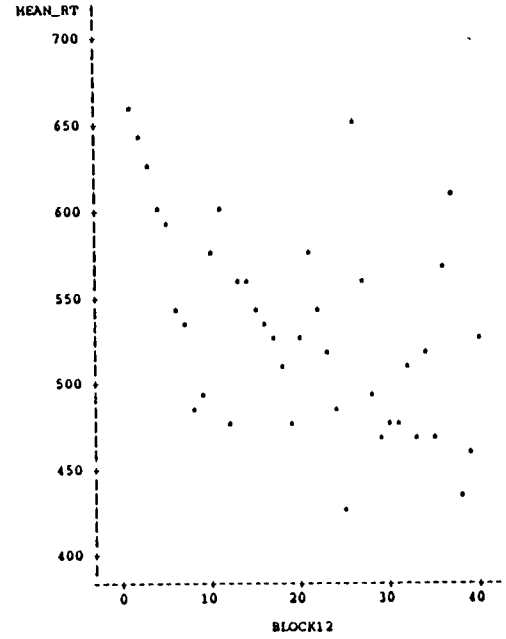
---------- COND=2 PRACSET=2 SUBJECT=7 ----------

Plot of MEAN_RT*BLOCK12. Symbol used is '*'.



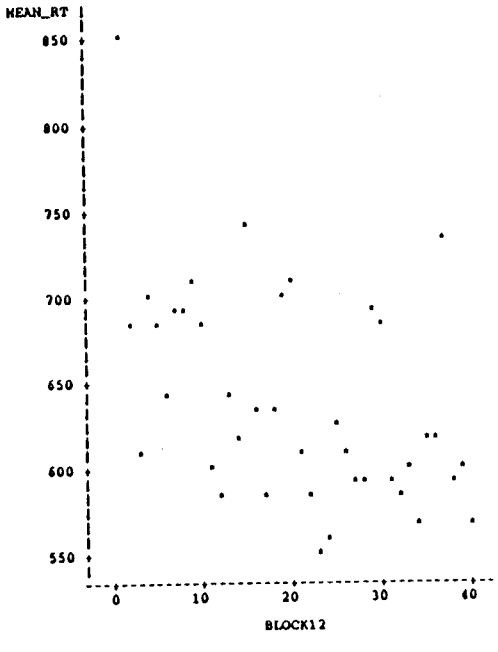

---------- COND=2 PRACSET=2 SUBJECT=8 ----------

Plot of MEAN_RT*BLOCK12. Symbol used is '*'.



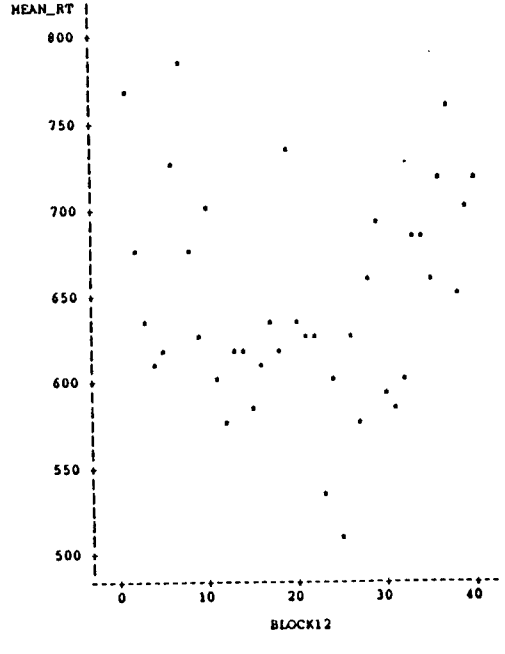
---------- COND=2 PRACSET=2 SUBJECT=15 ----------

Plot of MEAN_RT*BLOCK12. Symbol used is '*'.




---------- COND=2 PRACSET=2 SUBJECT=16 ----------

Plot of MEAN_RT*BLOCK12. Symbol used is '*'.



# APPENDIX C
# SPECIFICATIONS OF RED-GREEN-PURPLE COLOR SHADES USED IN EXPERIMENT 3

Experiment 3 used two different shades of red, of green, and of purple. The shades referred to as red1, green1, and purple1 were identical to the red-green-purple set used in Clawson et al. (in press) and in Experiments 1 and 2. The shades red2, green2, and purple2 were chosen to have different hues and luminances from the original set, under the constraints that they be consistently identified as red, green, and purple (respectively), and they be discriminable from the original same-name shades.

CIE chromaticity coordinates and luminance of both sets of shades are listed in Table C-1 (both 1931 and 1976 CIE coordinates; see Wyszecki & Stiles, 1982, for a description of these color spaces). Figure C-1 displays the CIE 1976 coordinates of the shades; in CIE 1976 coordinates, straight-line distances between points indicate degree of perceptual differences in chromaticity. The chromaticity coordinates and luminance were measured *in situ* with a spectroradiometer/photometer (Photo Research, Model PR703-A).

Table C-1
Specifications of the red-green-purple color shade sets used in Experiment 3: CIE 1931 coordinates (x, y), CIE 1976 coordinates (u', v'), and luminance (in candelas per meter squared).

| Shade Name | x | y | u' | v' | luminance |
|---|---|---|---|---|---|
| green1 | 0.3310 | 0.5780 | 0.1428 | 0.5609 | 50.48 |
| green2 | 0.3890 | 0.5305 | 0.1812 | 0.5559 | 45.70 |
| purple1 | 0.2601 | 0.1433 | 0.2477 | 0.3072 | 18.93 |
| purple2 | 0.3231 | 0.1966 | 0.2742 | 0.3754 | 11.74 |
| red1 | 0.5743 | 0.3652 | 0.3685 | 0.5273 | 12.88 |
| red2 | 0.5933 | 0.3489 | 0.3955 | 0.5233 | 27.63 |

To demonstrate that the new shades were consistently named red, green, and purple and that they were discriminable from the original shades, a separate group of subjects were asked to identify and discriminate between the shades. Subjects were nine psychology graduate students and a psychology postdoctoral fellow. The subjects made up two groups of five subjects each. One group of five subjects viewed green1 and green2 along with red1, purple1, one other potential shade of green, and two potential shades of red and of purple that were not used in Experiment 3. The other group viewed red1, red2, purple1, purple2, along with one other potential shade of purple and three other potential reds that were not used in Experiment 3. All colors were displayed on the computer screen as the nonword "wob" which is similar in length and shape to the shortest color word used (red) in Experiment 3.

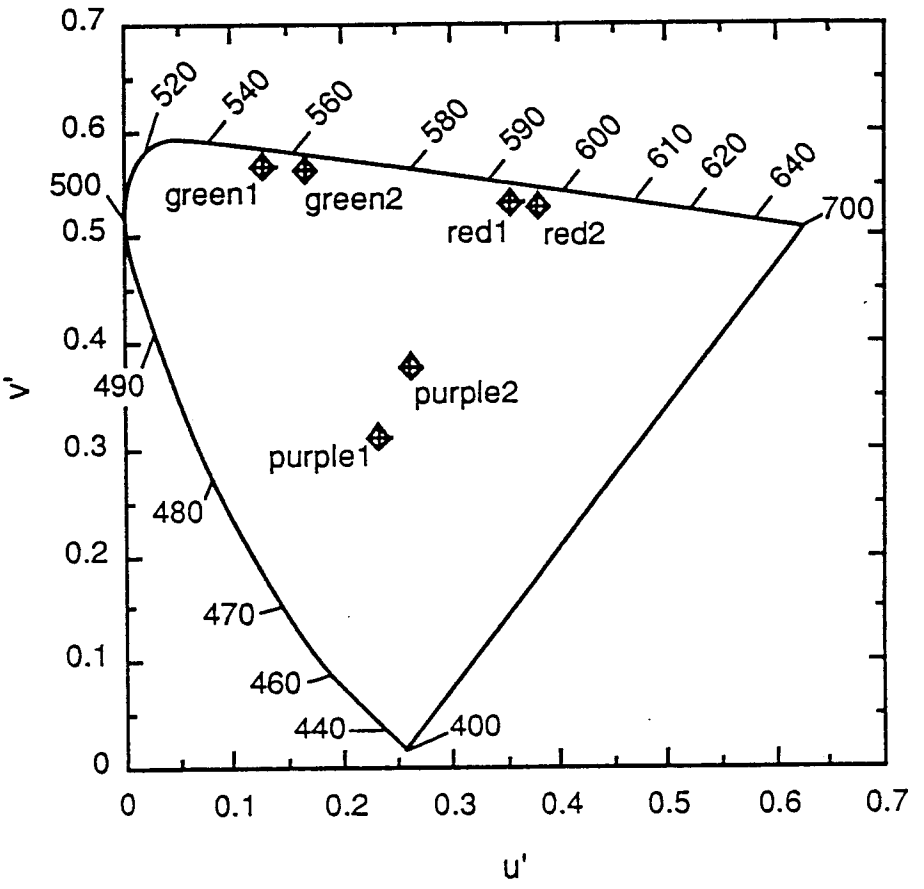



Figure C-1 . CIE 1976 coordinates (u', v') of the red, green, and purple shades used in Experiment 3

Each group completed two tasks on their colors, first completing an identification task, then completing a discrimination task. The identification task consisted of three blocks, each of which presented all of the colors once in pseudorandom order; subjects in the first group therefore had a total of 27 trials and subjects in the second group had a total of 24 trials. For each trial, the stimulus appeared in the center of the screen for 700 ms, mimicking the length of time that a typical stimulus would remain on the screen in Experiment 3 (because in that experiment, the stimuli disappeared when the subject responded). Subjects were instructed that after every stimulus they were simply to write on a numbered piece of paper the name of color they had just seen, followed by a number in the range 1-3 indicating how good an example of that named color the stimulus was. Subjects were instructed that a rating of "1" meant the stimulus was a good example of the named color, "2" meant it was an "OK" example, and "3" meant it was a poor example; the meanings of the ratings were also written in the corner of the numbered answer sheet. Subjects who needed further instruction on the ratings were shown examples of a poor blue and a good blue that were part of the computer case. After a subject had written a color name and rating for one stimulus, the subject pressed the c-key on the computer keyboard in order to be shown the next stimulus, and so on until all trials had been completed.

After the identification task, subjects began the discrimination task. In the discrimination task subjects were presented with pairs of "wob" stimuli, pairs made up either of a new shade twice or of a new shade and its corresponding original shade (pairs made up of an original shade twice were not included in an effort to reduce fatigue). There were two blocks, each of which included two trials on every new-new pair and two trials in each of the two possible orders for every new-original pair. Subjects in the first group therefore saw 54 trials, and subjects in the second group had 48 trials. For each trial the first stimulus appeared on the screen for 700 ms, followed by a black screen for 1,000 ms, followed by the second stimulus of the pair for 700 ms, mimicking the planned exposure times and intertrial interval of Experiment 3. The task was then to respond "same" (by pressing the m key) or "different" (by pressing the z key); the key assignments were also written in the corner of the subjects' identification task response sheets. The computer recorded for each trial the accuracy of the subject's response. After the subject had pressed one of the keys, the computer prompted the subject to press the space bar for the next pair. When the space bar was pressed, the screen went black for 1,000 ms then showed the first stimulus in the next pair. This procedure continued until all trials had been completed.

The identification and discrimination data for green1 and green2 from the first group and for red1, red2, purple1, and purple2

from the second group are included here. Both green1 and green2 were identified as "green" by all subjects on all presented trials. Red1 was identified by one subject as "dark red" on all trials, but as "red" by the remaining four subjects on all trials; red2 was identified as "red" by all subjects on all trials. Purple1 was identified as "purple" by all subjects on all trials; purple2 was identified as "deep purple" by one subject on one trial, but on the two other trials and for all other subjects it was identified as "purple." As well as the naming data, average ratings across the three blocks for all subjects are available for the reds and purples; the ratings data for the greens are no longer available. Red1 was rated an "OK" example of red (mean rating = 1.83), with the one subject who identified red1 as "dark red" giving it an average rating of 1.00. Red2 was rated between "good" and "OK" (M = 1.40). Purple1 was rated "good" (M = 1.00), and purple2 was rated as an "OK" purple (M = 1.72), with a rating of 2 the one time it was identified as "deep purple." Thus, all shades used in Experiment 3 were identified consistently with their intended color names.

The discrimination data are displayed in Table C-2 which shows proportion correct, across all subjects on all trials, for the new-new pairs (on which the correct response was "same") and for the new-original pairs in either order (on which the correct response was "different"). Subjects consistently discriminated between the new and original shades of all three colors.

Table C-2
Proportion correct on the discrimination task across all subjects on all trials for each same (new shade twice) and different (new shade and original shade) pair of stimuli.

| New shade | On same pairs | On different pairs |
|---|---|---|
| red2 | .85 | 1.00 |
| green2 | .94 | .97 |
| purple2 | .90 | .85 |